

(12)特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2004年10月21日 (21.10.2004)

PCT

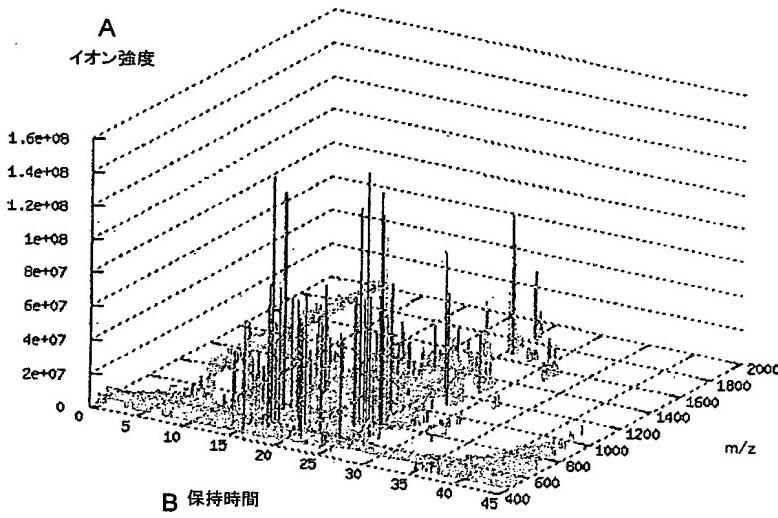
(10) 国際公開番号
WO 2004/090526 A1

- | | | |
|--|----------------------------|--|
| (51) 国際特許分類 ⁷ : | G01N 27/62, 30/72 | (72) 発明者; および |
| (21) 国際出願番号: | PCT/JP2004/004621 | (75) 発明者/出願人(米国についてのみ): 萩原淳 (OGIWARA, Atsushi) [JP/JP]; 〒1930942 東京都八王子市柄田町 521-6 マイキヤッスル柄田 310 号 Tokyo (JP). 川上 隆雄 (KAWAKAMI, Takao) [JP/JP]; 〒3360022 埼玉県さいたま市南区白幡 4-19-2 吉野マンション 303 号 Saitama (JP). 西村 俊秀 (NISHIMURA, Toshihide) [JP/JP]; 〒3001243 茨城県つくば市池向 1512-45 Ibaraki (JP). |
| (22) 国際出願日: | 2004年3月31日 (31.03.2004) | |
| (25) 国際出願の言語: | 日本語 | |
| (26) 国際公開の言語: | 日本語 | |
| (30) 優先権データ:
特願2003-095732 | 2003年3月31日 (31.03.2003) JP | (74) 代理人: 平木 祐輔, 外 (HIRAKI, Yusuke et al.); 〒1050001 東京都港区虎ノ門一丁目17番1号 虎ノ門5森ビル 3階 Tokyo (JP). |
| (71) 出願人(米国を除く全ての指定国について): 株式会社メディカル・プロテオスコープ (MEDICAL PROTEOSCOPE CO., LTD.) [JP/JP]; 〒1630207 東京都新宿区西新宿 2-6-1 新宿住友ビル Tokyo (JP). | | (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, |

[続葉有]

(54) Title: SAMPLE ANALYZING METHOD AND SAMPLE ANALYZING PROGRAM

(54) 発明の名称: 試料解析方法及び試料解析プログラム



A...ION INTENSITY
B...HOLDING TIME

WO 2004/090526 A1

(57) Abstract: A sample analyzing method and a sample analyzing program for analyzing a component contained in a sample with an excellent analysis ability. The method comprises a step (a) of correcting at least a one-dimensional parameter in multi-dimensional data collected by analyzing a sample and a step (b) of comparing the corrected data obtained at step (a) with data on other samples.

(57) 要約: 試料に含まれる成分を分析するに際して、優れた分析能を達成することができる試料解析方法及び試料解析プログラムを提供する。本発明に係る試

[続葉有]



DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC,

NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 國際調査報告書
- 請求の範囲の補正の期限前の公開であり、補正書受領の際には再公開される。

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイドスノート」を参照。

明 細 書

試料解析方法及び試料解析プログラム

5 技術分野

本発明は、試料の分析の結果として得られた多次元データを用いた試料解析方法及び試料解析プログラムに関する。

背景技術

10 例えれば、液体クロマトグラフィー（以下 LC と略記する）と質量分析（以下 MS と略記する）を連結した液体クロマトグラフィー質量分析（以下 LC-MS と略記する）の結果として、横軸に質量／電荷比（以下 m/z と略記する）、縦軸にイオン強度をとったグラフとして 2 次元上にスペクトルデータを得ることができる。ここで、LC の役割は、MS の処理能力に適応させるために、試料を単に分画すること 15 にある。

すなわち、LC によって分画された試料を MS で分析することによって、上述したような 2 次元のスペクトルデータを得ることができ、試料中の成分分析を行うことができる。ところが、LC の役割を単に分画に限定することにより、従来の LC-MS では、検体中から検出・識別できるタンパク質の種類が網羅的でなく、分析 20 能・解析能が低いといった問題がある。

一方で、クロマトグラフィーを単に分画だけでなく試料の特性を現す情報として利用する目的で、複数のクロマトグラフィー結果を比較できるようにするために、時間軸を補正して重ね合わせる (align) 方法がいくつか提案されている。その代表的なものとして、Dynamic Time Warping (以下、DTW と略記する) および 25 Correlation Optimized Warping ((以下、COW と略記する) がある。いずれも動的計画法アルゴリズムに基づいた 1 つの実装形態として、それぞれユークリッド距離、あるいは相関を 2 つのクロマトグラフィーの距離もしくは類似性の指標とするものである (V. Pravdova, B. Walczak, D. L. Massart, "A comparison of two algorithms for warping of analytical signals", Anal. Chim. Acta 456: 77-92

(2002))。しかしながら、これらの方法は、クロマトグラフィーの時間軸と信号強度との2次元で表されるクロマトグラムに適用されるものであるため、多次元データにおける少なくとも1次元のパラメータを補正するものではない。

さらに、このような重ね合わせ方法は、比較対象となるクロマトグラムあるいはスペクトログラムがある程度類似していることが前提となる。実際、DTWにせよCOWにせよ、比較対象のプロファイル間における距離の最小化あるいは相関の最大化を目指した重ねあわせを行うため、比較対象のプロファイルの共通性が乏しい場合には適切な重ねあわせが得られない可能性が十分にある。このように高い共通性を前提とした方法は、例えば、実際の疾患病態解析や薬剤応答性解析のように、多くの因子の変動が予想され、しかもそれぞれの変動量が微細で、個体差や測定誤差などに紛れそうな場合には不適切である。

そこで、本発明は、上述したような実状に鑑み、試料に含まれる成分を分析するに際して、優れた分析能を達成することができる試料解析方法及び試料解析プログラムを提供することを目的とする。

15

発明の開示

上述した目的を達成した本発明は以下を包含する。

(1) 試料の分析の結果として得られた多次元データにおける、少なくとも1次元のパラメータを補正する工程aと、上記工程aにより得られる補正後のデータを複数の試料について比較する工程bとを含む試料解析方法。

本試料解析方法において、多次元データとしては、クロマトグラフィー質量分析の結果として得られる、質量／電荷比を示すパラメータと、イオン強度を示すパラメータと、保持時間を示すパラメータとからなる3次元データを挙げることができる。また、このとき、工程aでは保持時間を示すパラメータを補正することが好ましい。

また、ここで、補正対象のパラメータを除いたパラメータに関するプロファイルを参照プロファイルと定義し、複数の試料間で複数の参照プロファイルに関する配置類似性の尺度となる評価関数を与えることができる。この場合、工程aでは、当該評価関数の値を最適化する最適解探索問題として各プロファイルの配置

を求めることができる。

このとき、評価関数は、以下の①～⑤からなる群から選ばれる 1 以上の項で定義されることが好ましい。

5 ①補正対象のパラメータに関するプロファイル間の類似性及び/又は距離に関する項

②参照プロファイルに関するプロファイル間の類似性及び/又は距離に関する項

③比較対象となるプロファイル間でのデータ点の一致の程度に関する項

④比較対象となるプロファイル間でのデータ点の不一致の程度に関する項

10 ⑤比較対象となるプロファイル間での標準物質由来シグナルの一致又は不一致の程度に関する項

⑥比較操作を繰り返す際に、過去の比較における一致の程度に関する項

また、工程 a では、補正対象のパラメータについて最適解探索問題として評価関数の値を最適化する際に動的計画法のアルゴリズムを用いることができる。この動的計画法のアルゴリズムでは、補正の対象となるパラメータに含まれるデータ点の最適な対応関係を、スコアを算出して評価する際に、標準物質に由来するデータ点に関する対応関係については、当該スコアを良くする設定とすることが好ましい。さらに、この場合、標準物質に由来するデータ点に関する対応関係については、必ず指定された点で対応することを拘束条件とすることが好ましい。

上記（1）に係る試料解析方法は、特に上記工程 aにおいて、あらかじめ添加された標準物質に由来する情報を用いることによって、解析の精度をより向上できるとともに、補正処理の能力をも向上させる。本発明に係る試料解析方法のなかでも、このような特徴を備える方法を *internal standard guided optimal profile alignment (i-OPAL)* 法と命名する。

また、上述した本発明に係る試料解析方法は、各種データを入力する機能を有する入力手段、プログラムに従って演算処理を実行する機能を有する演算処理手段、当該演算処理の結果等を表示する機能を有する表示手段を備えるコンピュータに実行させるためのプログラムとして実現することができる。

一方、本発明に係る試料解析方法は、異なった種類の試料の間で量の異なる物質を検出・同定することができる。具体的には、複数の試料において多次元デー

タとして、クロマトグラフィー質量分析の結果として得られる質量／電荷比を示すパラメータと、イオン強度を示すパラメータと、保持時間を示すパラメータとからなる3次元データを測定し、複数の試料間で3次元データを比較し、イオン強度が有意に異なるシグナルを検出・同定することができる。これらのシグナルの特性、即ち、質量／電荷比および保持時間に十分近接した特性を持つシグナルを生じさせた物質に対して、更なる解析を施すことで、当該物質を同定することが可能である。

さらに、検出・同定工程を、例えば疾患由来試料と、健常者あるいは健常組織由来試料とに適用することにより、疾患群と健常群で有意に異なる存在量を示す物質を検出・同定することができる。これによって同定された物質は、バイオマーカーとして利用することができる。バイオマーカー検出・同定結果を利用して、疾患の診断や治療法の選択へ利用することが可能である。

さらにまた、検出・同定工程を、例えば特定の治療法あるいは薬物への応答性の差異、あるいは副作用の差異を示す患者集団由来の試料へ適用することで、治療法／薬剤応答性あるいは副作用のマーカーとなる物質の検出・同定へ利用可能である。

本明細書は本願の優先権の基礎である日本国特許出願 2003-95732 号の明細書及び/又は図面に記載される内容を包含する。

20 図面の簡単な説明

図 1 は、本発明に係る試料解析方法及び試料解析プログラムにより取得された3次元のスペクトルデータの一例を示す図である。

図 2 は、3次元データの一例を示す図である。

図 3 は、図 2 に示した3次元データとの対応関係を検索するために設定したもの 25 う一つの3次元データの一例を示す図である。

図 4 は、図 2 に示した3次元データと図 3 に示した3次元データとの最適な対応配置を示した図である。

図 5 は、図 2 に示した3次元データと図 3 に示した3次元データとの最適な対応配置を検索する際の概念を示す図である。

図 6 は、図 5 に示した最適配置探索において、標準物質由来の情報を利用して経路を限定すると、探索空間のうちの灰色部分はもはや探索する必要がないことを示した図である。

図 7 は、標準物質由来の情報をより多く利用して経路の拘束条件を増やすと、
5 探索不要な空間が増えて、探索効率がいっそう向上することを示した図である。

図 8 は、本発明に係る試料解析プログラムによって、同一種類の 5 つの測定結果で時間軸上揺らいでいる波形を、1 つに重ね合わせてシグナルを足し合わせた結果を示した図である。

図 9 は、本発明に係る試料解析プログラムによって、異なる 7 種類の試料測定
10 から得られたプロファイルを時間軸をそろえて重ね合わせることにより、時点ごとに異なった試料間での相互比較が行えるようになったことを示した図である。

図 10 は、本発明に係る試料解析プログラムによって、異なった 2 種類の試料間の差スペクトルを計算して示した図である。

図 11 は、本発明に係る試料解析プログラムによって、試料群間に有意な量的変
15 動があるシグナルとして選別されたものを示した図である。

図 12 は、本発明に係る試料解析プログラムを、実際の臨床検体を用いたマーカー探索に応用し、異なった病理診断結果によるグルーピングに応じてシグナルを分類できることを示した図である。

図 13 は、図 12 に示した結果に対し、さらに統計検定を行うことで、異なった
20 病理診断結果に応じて量的に変化しているシグナルを拾い出したものを示した図である。

図 14 は、図 13 に示した結果の各シグナルに対し、さらに MS/MS による蛋白質同定との関連付けを行った結果を示した図である。

図 15 は、図 14 に示したような形で既知の蛋白質と関連付けられたもののうち、
25 特に癌の転移に関連することが知られている蛋白について、本発明に係る試料解析システムで見出された結果の一部を示した表である。

発明を実施するための形態

以下、図面を参照して本発明を詳細に説明する。

1. 試料の調整

本発明に係る試料解析方法では、先ず、解析対象の試料を採取する。解析対象の試料としては、特に限定されないが、例えば、動物個体由来の臓器の組織切片や血漿・リンパ液などの体液成分、植物の緑葉や花弁などの器官、環境中の土壤・水成分などがあげられる。
5 これらの試料に含まれる分析対象物質としては、特に以下に限定されないが、例えば、有機化合物、無機化合物、有機金属化合物、金属イオン、ペプチド、蛋白質、金属蛋白質、リン酸化を含む翻訳後修飾を受けたペプチド、リン酸化を含む翻訳後修飾を受けた蛋白質、核酸、糖質、脂質などがあげられるが、特に望ましくはペプチド、蛋白質、金属蛋白質、翻訳後修飾を受けたペプチドもしくは蛋白質である。
10

また、採取した試料は、分析の目的及び採取した試料の特性に合わせて、必要であれば各種処理を施すことが好ましい。例えば、(ア) タンパク質群の分離あるいは分画、(イ) タンパク質群の酵素的及び／又は化学的切断、(ウ) 切断によって生じたペプチド混合物の分離あるいは分画、および(エ) 標準物質の添加、
15 全ての要素あるいは一部の要素の組み合わせて行う分析前調製を施すことが好ましい。

より具体的に、「(ア) タンパク質群の分離あるいは分画」は、一次元のドデシル硫酸ナトリウム (SDS) 電気泳動法、二次元電気泳動法、キャピラリー電気泳動法、イオン交換クロマトグラフィー、ゲルろ過クロマトグラフィー、順相クロマトグラフィー、逆相クロマトグラフィー、アフィニティーコロマトグラフィー、
20 或いはこれらの組み合わせによる多次元分離・分画等によって行うことができる。

また、「(イ) タンパク質群の酵素的及び／又は化学的切断」は、トリプシン消化、キモトリプシン消化、Lys-C 消化、Asp-N 消化、Glu-C 消化、臭化シアン分解
25 或いはこれらの組み合わせによる切断等によって行うことができる。

さらに、「(ウ) 切断によって生じたペプチド混合物の分離あるいは分画」は、一次元のドデシル硫酸ナトリウム (SDS) 電気泳動法、二次元電気泳動法、キャピラリー電気泳動法、イオン交換クロマトグラフィー、ゲルろ過クロマトグラフィー、順相クロマトグラフィー、逆相クロマトグラフィー、アフィニティーコロマトグラフィー或いはこれらの組み合わせによる多次元分離・分画等によって行う

ことができる。

- さらにまた、「(エ) 標準物質の添加」において標準物質は、選択したイオン化法にてイオン化できるものであって、測定の LC 保持時間の範囲内に溶出するものであって、溶出時間および分子イオン強度の再現性が高いものを選択することが
5 好ましい。このような好ましい標準物質としては、例えば、有機化合物、無機化合物、有機金属化合物、金属イオン、ペプチド、蛋白質、金属蛋白質、リン酸化を含む翻訳後修飾を受けたペプチド、リン酸化を含む翻訳後修飾を含む蛋白質、核酸、糖質、脂質など、より好ましくは、ペプチド・蛋白質で市販品、天然に存在する物質あるいは合成された物質を挙げることができる。
10 以上、(ア) ~ (エ) に示した分析前における各種処理は、例えば、「ア、エ、イ、ウの順」、「エ、イ、ウの順」、「イ、エ、ウの順」、「エ、アの順」、「エ、イの順」「イ、エの順」又は「エのみ」で行うことができる。

2. 試料分析

- 15 次に、試料を分析することによって、試料に関する多次元データを取得する。具体的には、試料を LC-MS により分析し、 m/z 、イオン強度及び保持時間からなる多次元データを測定する。ここで、LC-MS により分析するとは、試料をクロマトグラフィーの原理に従って分離又は分画し、その後、分離又は分画された試料に含まれる成分を質量分析の原理で測定することを意味する。なお、保持時間は、
20 試料をクロマトグラフィーの原理に従って分離又は分画する際の時間として測定される。また、 m/z 及びイオン強度は、質量分析の結果として測定される。

また、クロマトグラフィーの原理としては、特に限定されないが、逆相クロマトグラフィー、キャピラリー電気泳動、アフィニティーコロマトグラフィー、クロマトフォーカシング、等電点フォーカシング、ゲルろ過クロマトグラフィー等の各種クロマトグラフィーの原理を適用することができる。特に、本明細書において LC と表記する場合、液体クロマトグラフィーのみを意味するのではなく、広く一般的なクロマトグラフィーを意味する。

LC-MS におけるクロマトグラフィーでは、再現性の高い溶出プロファイルが得られること、分離能が高いため、適当なイオン化のインターフェースを介して MS

に直接分子イオンを導入することが可能であることが好ましい。

より具体的に、液体クロマトグラフィーにおける好ましい条件としては次の通りである。試料中のペプチド群を分析対称とする場合には、水・アセトニトリル溶液に低濃度の蟻酸などの強酸を含む溶離液を用いた C18 カラムによる逆相液体クロマトグラフィーが好ましい。また蛋白質群を分析対称とする場合には、水・アセトニトリル溶液に低濃度の蟻酸などの強酸を含む溶離液を用いた C4 カラムによる逆相液体クロマトグラフィーが好ましい。
5

質量分析は、特に限定されないが、磁場型質量分析計、飛行時間型質量分析計、四重極質量分析計、イオントラップ質量分析計、フーリエ変換質量分析計または 10 これらのハイブリッド及びタンデム質量分析計等により行うことができる。より好ましくは、エレクトロスプレーイオン化またはナノエレクトロスプレーイオン化と結合できる磁場型質量分析計、飛行時間型質量分析計、四重極質量分析計、イオントラップ質量分析計、フーリエ変換質量分析計またはこれらのハイブリッド若しくはタンデム質量分析計を用いて質量分析を行うことが好ましい。
15

LC-MS における質量分析では、再現性の高い質量スペクトルが得られること、 500ppm 以下の高い質量精度を有すること、一定範囲の m/z の分子イオンに対して衝突誘起解離 (CID) をかけ、当該分子イオンのフラグメントイオンの質量スペクトルが得られることが好ましい。

このように、試料を LC-MS により分析し、 m/z 、イオン強度及び保持時間を測定することによって、試料の分析結果を 3 次元データとして取得することができる。なお、LC-MS による分析は、保持時間に関するデータ、 m/z に関するシグナル及びイオン強度に関するデータを、入力手段を介してコンピュータに入力し、詳細を後述するアルゴリズムに従って演算処理手段によって処理を行うことにより図 1 に示すような 3 次元データとして取得することができる。本アルゴリズム 25 は、コンピュータソフトウェアに搭載することができる。当該ソフトウェアをコンピュータにインストールすることによって、CPU 等の演算処理手段により本アルゴリズムをコンピュータ上で実現することができる。従って、図 1 に示すような 3 次元データは、コンピュータの表示装置に表示することができる。

従前の LC-MS による解析方法においては、単に試料の分画のために LC を行つ

ているため、保持時間は解析対象パラメータとして使われておらず、試料の分析結果として横軸に m/z 、縦軸にイオン強度をとった 2 次元データが解析対象となりうるに過ぎなかった。これに対して本発明に係る解析方法によれば、試料の分析結果を 3 次元空間上にプロットされるプロファイルとして取得することができるため、試料の分析能を飛躍的に向上させることができる。具体的には、本発明に係る解析方法によれば、保持時間を示す軸の方向に広がりを持った多数のスペクトルの重ね合わせとしてデータを取得でき、従前の解析方法と比較してより多数の成分に関して、スペクトルに基づく同定を行うことができる。このため、例えば、複数の試料について得られた多次元データを比較することで各試料の成分分析をより厳密に行うことができる。

3. データ解析

次に、本発明に係る解析方法においては、以上のように測定した保持時間を、演算処理手段の制御のもとに本発明に係るアルゴリズムによって補正することもできる。ここで、一般に保持時間は、LC における移動相の組成、流速、カラム温度等のファクターが時間的に微小な変化を生じることから、非線形的に変動することが多い。したがって、本発明に係る解析方法で取得された 3 次元データに関しても、複数の試料について解析を行った場合に試料間の保持時間を示す軸が非線形的に変動していることが考えられる。そこで、本発明に係るアルゴリズムにおいては、保持時間の補正（以下、時間軸補正とも言う）を行う。

但し、本発明に係るアルゴリズムが対象とする時間軸補正是、従前のことと DTW アルゴリズム等によるクロマトグラムの時間軸補正のような、保持時間—信号強度の 2 次元空間で表現される單一次元プロファイルの補正ではない。本発明が対象とするデータは、時間軸に対して補正されるべきプロファイルが少なくとも 2 次元以上の多次元で表現されるものである。

以下、当該アルゴリズムについて説明するが、当該アルゴリズムは保持時間の補正に限定されず、多次元のパラメータが得られた場合に少なくとも 1 次元のパラメータを補正する場合に広く適用することができる。言い換えると、当該アルゴリズムは、試料の分析の結果として得られた多次元のパラメータ（例えば 3 次

元のパラメータ) における、少なくとも 1 次元のパラメータを補正する際に適用することができる。従って、以下の説明においては、 $p+q$ 次元の測定データを取得した場合のアルゴリズムについて説明する。

先ず、補正対象とするパラメータを含む p 次元の測定値を $(x_1 \dots x_p)$ とし、補正の際に参照する q 次元の測定値を $(y_1 \dots y_q)$ とすると、データの集合 (プロファイル) z は $z = (x_1 \dots x_p \ y_1 \dots y_q)$ となる。ここで、 x 及び y は、データ点の個数 N の次元を持つ列ベクトルである。

なお、データ点とは、上記プロファイル行列 (z) の 1 つの行を構成する $p+q$ 次元のベクトルであり、測定対象の 1 つの要素について、測定パラメータと値の組を表している。特に、 $n \in \{1, \dots, N\}$ 番目のデータ点を $\overline{Z(n)} = (x_1(n) \dots x_p(n) \ y_1(n) \dots y_q(n))$ のようにも表す。

また、補正の基準となる測定値を $\overline{Z(*s)} = (x_1(*s) \dots x_p(*s) \ y_1(*s) \dots y_q(*s))$ とする。ここで、 s は $1 \leq s \leq S$ (S は基準点の数) を意味する。また、 $\overline{Z(*s)}$ は、いずれも各基準点のとる値が推定可能な範囲に収まらなくてはならない。

さらに、本アルゴリズムにおいて補正を行うためには 2 つ以上のプロファイルデータ $Z^{(1)} = (x_1^{(1)} \dots x_p^{(1)} \ y_1^{(1)} \dots y_q^{(1)})$ 及び $Z^{(2)} = (x_1^{(2)} \dots x_p^{(2)} \ y_1^{(2)} \dots y_q^{(2)})$ が必要となる。

以上のような定義の下で、本アルゴリズムにおいては先ず、 p 個のパラメータ軸 $x_1 \dots x_p$ それぞれにおいて取りうる値を量子化する。但し、量子化プロセスは、計算精度と計算時間との兼ね合いで行うものであり、値のとりうる場合の数が十分計算可能な範囲にあるならば、必ずしもこの段階で行わなくてもよい。次いで、 p 個のパラメータ軸 $x_1 \dots x_p$ それぞれにおいて $x_i^{(1)}$ 及び $x_i^{(2)}$ (但し $i \in \{1, \dots, p\}$) の各データ点を、順列を保って対応付けを行う。なお、一般に、 $Z^{(1)}$ 及び $Z^{(2)}$ に含まれるデータ点の個数は異なりうるので、全てのデータ点が 1 対 1 に対応するわけではなく、対応する相手のないデータ点も含むことに留意する。

このとき、例えば、以下のような評価関数を用いてプロファイル全体での対応付けの評価得点 E を算出する。なおこの評価得点は、類似性を表す尺度として、大きいほどよい「得点」として定義することも、逆に距離を表す尺度として、小さければよい「失点」として定義することも可能である。以下では失点としての

定義で説明する。

$$E = \sum_{i=1, j=1}^{N_1, N_2} f(x'_1, x'_1, \dots, x'_p, x'_p, y'_1, y'_1, \dots, y'_q, y'_q)$$

ここで x'_r は、 i 番目のデータ点における r 番目のパラメータの値を表し、 N_1, N_2

はそれぞれ 1 番目と 2 番目のプロファイルにおけるデータ点の総数である。また、

5 上記評価関数において関数 f は、対応する点の類似度合いの距離を与える関数で、
例えば以下のような関数を挙げることができる。

$$\begin{aligned} & f(x'_1, x'_1, \dots, x'_p, x'_p, y'_1, y'_1, \dots, y'_q, y'_q) \\ &= \sum_{r=1}^p \alpha_r \cdot \delta_r(i, j) \cdot \|x'_r - x'_r\| + \sum_{s=1}^q \beta_s \cdot \|y'_s - y'_s\| - \sigma \prod_{r=1}^p \delta_r(i, j) + \sum_{r=1}^p \pi_r \cdot \overline{\delta_r(i, j)} - \theta(i, j) \end{aligned} \quad \cdots (I)$$

ここで上記式(I)における、右辺第 1 項目は補正すべきパラメータ x_r において当該パラメータの測度における差異の程度に応じたペナルティ、第 2 項目は合わ

10 れるべき測定パラメータ y_s が補正後どれだけずれたかを当該パラメータ測度上の距離に応じたペナルティ、第 3 項目はパラメータ補正によって 2 つの点が全てのパラメータにおいて一致したことのボーナスとして与えられるスコア、逆に第 4 項目は補正対象パラメータ軸上での 2 つの点が一致しなかったことによるペナルティスコアに相当する。また第 5 項目は、後述するように、標準物質によるシ

15 グナルの一致をボーナスとして評価するための項である。

また、上記式 (I) における、 α 、 β 、 σ 及び π は、それぞれを含む項における係数であり、適宜設定できる値である。一例としては、 α を 1.0 とし、 β を 0.1 とし、パラメータ補正によって点が一致した場合には σ を 0 とし、一致しなかった場合には π を 100 とすることができる。

20 なお、関数 $\delta_r(i, j)$ は、着目するパラメータ r の値が i, j によって指定されるデータ点において対応した場合は 1、しなかった場合は 0 を与える関数、逆に $\overline{\delta_r(i, j)}$ は対応した場合が 0、非対応の場合が 1 となる関数である。

上記式(I)において、第 2 項目は、補正対称のパラメータを除いたパラメータに関するプロファイル（参照プロファイル）について、試料間での配置類似性の尺度を示している。

また、ここでは 2 点の不一致によるペナルティを与える式として、対応・非対

応によって定数となる例を示したが、所定の関数によって算出される値であっても良い。例えば、隣接するデータ点が対応するか否か、対応しないデータ点が出現した列の長さ等を考慮した関数によって第4項目を算出することができる。

また、上記式(I)において、ノルム $\|x\|$ は一般的なベクトル空間上の距離を表し、

- 5 必ずしもユークリッド距離に限定するものではない。さらに $\|y_s^i - y_s^j\|$ のように2点間の値の差を計算する際に、対応する点がない場合は、値を0(または適当な欠損値の代替値)に置き換えて計算するものとする。

なお、本発明において評価関数は、上記式(I)に示す関数に限定されるものではない。例えば、このようなデータ点 (i, j) 間の補正対象パラメータ或いは参照パラ
10 メータの距離の線形結合だけではなく、両者の距離に応じた任意の関数、更には直前若しくはそれまでに連続して対応したデータ点列におけるパラメータ間の距離も勘案した関数を定義することも可能である。また、評価関数としては、上記式(I)に限定されず、試料間において参照プロファイの配置類似性の尺度となるような関数を定義することも可能である。

15 あるいは、ここでは失点としての例を示したが、上記式(I)における右辺各項の符号を逆転させた上で、距離の部分を相関などに置き換えることにより、逆に類似性を表す指標として、大きければ大きいほどよい得点としての評価関数を定義することも容易に可能であり、そのような評価関数も本アルゴリズムに適用することができる。

20 上記式(I)の第5項で例示したように、対応する点が標準物質由来の基準点か否かで、例えば次のような特別な得点を与えるものとする。即ち、対応するデータ点がいずれも標準物質由来のものであった場合は、 $\theta(i, j) = S_m$ として、はるかに大きな得点を設定することで、評価関数(この場合は距離即ち失点としての定義)が大きな負の値をとりうることで結果的にこのような対応関係が望ましいことを
25 定義する。さらに、一方が標準物質由来であるのに他方がそうでない場合は、逆に $\theta(i, j) = -S_d$ としてはるかに大きな距離となるように定義することも可能である。

以上で説明した評価関数を最適化するアルゴリズムによって、上記「2. 試料分析」で取得した3次元データに関して、保持時間を示すパラメータの補正を行

うことができる。上記「2. 試料分析」で取得した3次元データに関して、最適化アルゴリズムを適用する場合、以下の(a)～(d)の手順に従って説明することができる。

(a) 保持時間補正の概念

5 保持時間を補正する操作は、 m/z 、イオン強度及び補正時間からなる单一の三次元パラメータ集合体を対象とするのではなく、2つの三次元パラメータ集合体の比較によって実現される。三次元パラメータ集合体は、図2に示すように、 m/z と保持時間とをそれぞれ行と列にとった行列において、 m/z 及び保持時間が対応する位置の行列要素にイオン強度が入るような形で表される。保持時間を補
10 正する対象の三次元パラメータ集合体をそれぞれ $Z^{(1)}$ 及び $Z^{(2)}$ とすると、保持時間の補正操作は、 $Z^{(1)}$ 及び $Z^{(2)}$ における2つの行列で、保持時間軸に相当する列の対応関係を決める操作（以下、「対応配置の検索」と呼ぶ）に他ならない。例えば、図2に示す行列を $Z^{(1)}$ の行列とし、図3に示す行列を $Z^{(2)}$ の行列とすると、図4のような配置が望ましい対応配置（重ね合わせ配置）である。

15 (b) 2つの3次元データにおけるパラメータ集合体間の対応配置の探索

図4に示すような対応配置を探索するためには、可能なすべての保持時間の対応付けを考える。この際、配置の対応関係の良し悪しを評価するスコアを定義し、配置毎にスコアを計算し、その中でもっともスコアがよいものを採用することで目的とする最適な対応配置を得ることができる。図5は、図2と図3に示した三次元パラメータ集合体 $Z^{(1)}$ 及び $Z^{(2)}$ に関して、保持時間の可能なすべての対応付けを示したものである。横方向に $Z^{(1)}$ の保持時間、縦方向に $Z^{(2)}$ の保持時間が記されているが、(イ) $Z^{(1)}$ 及び $Z^{(2)}$ でそれぞれ対応する保持時間のある場合が斜線、(ロ) $Z^{(1)}$ の所定の保持時間に対して $Z^{(2)}$ の方に対応するものがない場合が横線、(ハ) $Z^{(2)}$ の所定の保持時間に対して $Z^{(1)}$ の方に対応するものがない場合が縦線で示さ
20 れている。 $Z^{(1)}$ 及び $Z^{(2)}$ の全体的な保持時間の対応付けは、図5の格子の最左上角から最右下角にいたる経路を、これら斜線・横線・縦線をなぞることで求めることに相当する。但し、一度下がったり右に進んだら、進んだ点から逆に上や左に戻るような経路は許されない。なお、図5において太線で示された経路は、図4の対応配置に相当する。

(c) 保持時間対応配置の良否を判断するためのスコア

保持時間に関する対応配置の良否を判断するスコアは、例えば、次のようにして定義することができる。

i) 最左上点におけるスコア、すなわちまだ対応関係がまったく決まっていない点でのスコアを 0 とする。

ii) 前述の (イ) (ロ) 及び (ハ) のうちいずれかの場合をとることにより、対応関係が 1 段階進んだ場合は、その直前のスコアに対して、(イ) (ロ) 及び (ハ) 毎に決められたスコアを加算することで、新たな対応関係の点におけるスコアとなる。例えば、以下のように (イ) (ロ) 及び (ハ) 每にスコアを設定することができる。

(イ) の場合 (図 5において斜線方向に進む場合) :

この場合、所定の保持時間に関して、 $Z^{(1)}$ 及び $Z^{(2)}$ が互いに対応付けられるわけである。したがって、この場合、加算されるスコアとしては、 m/z パラメータとイオン強度パラメータが $Z^{(1)}$ 及び $Z^{(2)}$ 間でどれだけ類似或いは離れているかを反映した値を設定することができる。以下の説明では類似度としてスコアを定義した場合について説明する。例えば、 $Z^{(1)}$ において所定の m/z の元でイオン強度が検出されているのに $Z^{(2)}$ には同 m/z の元でイオン強度が検出されなかったケース、あるいはその逆のケースであれば、一定の値 (ペナルティスコア) を減じるようにスコアを設定することができる。また、所定の m/z において $Z^{(1)}$ 及び $Z^{(2)}$ それぞれにイオン強度が得られている場合、例えば両イオン強度の差の絶対値に所定の係数を乗じて算出される値 (ペナルティスコア) を減じるようにスコアを設定することができる。さらに、スコアとしては、両イオン強度の違いが大きければ大きいほど、得点が小さくなるような関数で算出されるものであっても良い。

一方、 $Z^{(1)}$ 及び $Z^{(2)}$ における保持時間のずれもスコアに反映させることもできる。例えば、 $Z^{(1)}$ 及び $Z^{(2)}$ における保持時間の差の絶対値に所定の係数を乗じて算出された値 (ペナルティスコア) を減じるようにスコアを設定することができる。スコアとしては、 $Z^{(1)}$ 及び $Z^{(2)}$ における保持時間の違いが大きければ大きいほど、得点が小さくなるような関数で算出されるものであっても良い。

なお、標準物質由来のシグナルが $Z^{(1)}$ 及び $Z^{(2)}$ で対応する場合には、後述する計

算方法の工夫に加えて、スコア算定上も特別な措置を施すことが好ましい。特に、これらの点が $Z^{(1)}$ 及び $Z^{(2)}$ 間で一致することは強く求められるので、 $Z^{(1)}$ 及び $Z^{(2)}$ とともに標準物質由来シグナルとして対応付けられる場合には大きな得点を、逆に一方だけしか標準物質由来シグナルが見つからない場合には大きな失点を与える。

5 (ロ) 及び (ハ) の場合 (図 5において縦又は横方向に進む場合) :

この場合、所定の保持時間に関して、 $Z^{(1)}$ 及び $Z^{(2)}$ において対応する保持時間を見出せなかったわけである。したがって、この場合、所定の値 (ペナルティスコア) を減じるようなスコアを設定する。

iii) このようにして図 5 の格子の最左上角から最右下角にいたるまで、段階的に 10 スコアを求めてゆき、最後に最右下角までいたった時点でのスコアが、得られた対応配置に対応するスコアになる。

(d) 保持時間に関する最適な対応配置を求める手順

基本的には、可能なすべての対応配置を列挙し、それぞれについてスコアを計算し、その中で最大のスコアを示す対応配置を選択すればよいわけであるが、上 15 述したように、スコアは漸化式で与えられるため、「動的計画法」に適している問題である。すなわち、3 次元データ $Z^{(1)}$ に含まれる i 番目の保持時間と $Z^{(2)}$ に含まれる j 番目の保持時間の対応関係を考える際には、(イ) $Z^{(1)}$ に含まれる $i-1$ 番目及び $Z^{(2)}$ に含まれる $j-1$ 番目に次いで $Z^{(1)}$ 及び $Z^{(2)}$ 両者ともに対応付けられる場合、(ロ) $Z^{(1)}$ に含まれる $i-1$ 番目及び $Z^{(2)}$ に含まれる j 番目に次いで $Z^{(1)}$ の保 20 持時間に対応する $Z^{(2)}$ のパラメータがない場合、(ハ) $Z^{(1)}$ に含まれる i 番目と $Z^{(2)}$ に含まれる $j-1$ 番目に次いで $Z^{(2)}$ の保持時間に対応する $Z^{(1)}$ のパラメータがない場合、の 3 通りを考えることとなる。いずれの場合も 1 段階前の状態におけるスコアがわかっていてれば $Z^{(1)}$ 及び $Z^{(2)}$ の (i, j) 番目のスコアを算出することが可能となる。

そこで (イ) (ロ) 及び (ハ) の 3 通りのうち、最もよいスコアを与えた場合のスコアとそこに至るまでの経路のみを記録しておき、このステップを図 5 に示した格子の最左上角の出発点から最右下角のゴールに達するまで続ける。そして、記録した経路を最右下角から出発点まで、逆にたどることにより、最適経路、すなわち $Z^{(1)}$ 及び $Z^{(2)}$ における保持時間に関して、最適な対応配置を求めることがで

きる。

以上、(a)～(d)の手順として説明した手法は、動的計画法に基づいた最適解探索手法と言い換えることができるが、本発明において適用可能なアルゴリズムは動的計画法に限定されるものではない。すなわち、目的の評価関数を最適化するより一般的な探索問題と捉えることで、他の最適探索アルゴリズムを用いて実装することも可能である。このようなアルゴリズムとしては、例えばA*アルゴリズム、遺伝的アルゴリズム(GA)、シミュレーテッドアニーリング(SA)、最急降下法等による非線形計画法等によって実装することもできる。

(a)～(d)の手順として説明した手法は、いわゆる動的計画法に基づく手法であり、動的計画法に基づいている点では、DTW法やCOW法と類似する部分がある。しかしながら、DTWやCOWは、評価関数の形や計算手法もユークリッド距離あるいは相関を評価関数として、時系列データ点列そのままあるいは一定間隔の時間区切りでセグメント化した上での区間ごとの比較として、さらに2つのプロファイルの時刻0を出発点とし、それぞれのプロファイルの終了時刻をゴールとする同じ大域的制約条件の下に探索を実施する方式に限定されたものである。さらに、DTWやCOWを用いる手法は、基本的に2次元データとして表される時系列プロファイル、即ち時間軸とシグナル強度軸で表されるようなデータセットに対し、時間軸の非線形伸縮を行うことによって強度のプロファイルを重ね合わせるものである。

したがって、DTWやCOWを用いる手法では、(1)特定の軸について特定の値をとる1つないし複数個の切断面を用いて、あるいは、(2)特定の軸に沿って全ての値を集約することによって、重ねあわせ操作を行わせることは、これらの方法の自然な拡張として容易に考えられる。例えば、DTWやCOWを用いる手法では、LC-MS分析で得られた保持時間、m/z、イオン強度からなる3次元データについても、いくつか特定のm/zに限定することにより、あるいは、全イオンクロマトグラム(TIC)のように保持時間軸に沿って全てのイオン強度を足し合わせることによって、時間軸補正が可能である。

しかしながら、(a)～(d)の手順として説明した手法では、DTWやCOWを拡張した方法とは異なり、補正対象の次元(保持時間軸)を除いた多次元プロファ

イルはそのままの形で直接比較することで、補正対象の次元を伸縮させることで、多次元プロファイルの重ねあわせを実現している。DTWやCOWを拡張した方法では、(1)特定の切片に限った手法をとると、精度を保ちつつプロファイル全体を重ね合わせるのと同じ結果が得られる保証がないこと、特定の切片を選択するための汎用性のある有効な手段がないこと、そのような保障なしに限定した重ねあわせ処理を行うことで、結果が恣意的になりうる危険があること、(2)TICのように情報を集約することで、せっかく次元を多くしてより分解能を向上させているメリットが得られなくなってしまう、といった問題がある。これに対して、(a)～(d)の手順として説明した手法では、上記(1)及び(2)の問題が全くなく、高精度にプロファイルの重ねあわせが可能となり、且つ、多次元データという高分解能を維持したままプロファイルの重ねあわせが可能となる。

4. データ解析における標準物質の役割

また、本発明に係る試料測定方法においては、以下に述べるような標準物質由来の情報を、本アルゴリズムを用いた計算手法に取り入れることで高い精度及び計算効率を達成することができる。

上記「2. 試料の分析」に先立って又は「2. 試料の分析」の途中で標準物質を入れることにより、測定及び解析の過程で生じる可能性のあるバイアスを補正すると同時に、これらの情報を用いることによって、上述した最適な対応配置、すなわち、プロファイル同士の重ねあわせをより正確にかつより効率的に行うことなどが可能となる。すなわち、標準物質を用いることによって以下のメリットを得ることができる。

- (1) あらかじめ既知の量の標準物質を入れることで、全体のシグナル強度を補正し、定量的な比較が可能となること
- (2) いくつかの標準物質を、補正すべきパラメータ（時間軸等）を合わせる際のランドマークとして使用できること
- (3) プロファイル重ねあわせにおいて、ある程度のプロファイル形状の共通性を与えることにより、重ねあわせが容易になること

これらメリットを上記「3. データの解析」において最大限発揮させるべく、

計算手法を以下のように改変することができる。すなわち、上述したアルゴリズムにおいて、標準物質に由来するシグナルのピーク部分を必ず通らなくてはいけない点として扱うように改変することができる。より詳しく説明すると、上述したアルゴリズムによる最適解探索では、図5に示したような格子状の探索空間上で、左上の出発点から右下の到着点に至る最適経路（太線）を探索するものであるが、いまここで仮に縦列上の保持時間15の点と、横の行で保持時間13の点がそれぞれ標準物質由来であったと仮定すると、解となりうる経路は必ずこの点を通らなければいけないというように拘束条件を設ける。そして、このような拘束条件を設定することによって、列15と行13を通る線で区分される探索空間のうちの左下と右上の部分空間を通る経路は排除されるため、探索しなければいけない空間を削減することができる（図6）。

このように、標準物質由来のシグナルについて拘束条件を設定するようにアルゴリズムを改変することによって、本発明に係る試料解析プログラムは、プロファイル同士の重ねあわせをより正確に行うことができるとともに、処理の能率を大幅に向上させることができる。

さらに、標準物質由来のシグナルの数が増えるにつれて、検索空間はより限定されることになるため、プロファイル同士の重ねあわせ精度はより向上するとともに、より能率向上が期待できる。実際、図7に示すように、丸印で示した点を標準物質由来シグナルの一一致点として拘束条件を設定すると、灰色でマスクした領域は探索空間から削除される。 n 種類の標準物質由来シグナルで時間軸が $n+1$ 個に分割されたとすると、最良のケースとして分割が等間隔で行われたならば、

探索空間は最大 $\frac{1}{n+1}$ まで減少する。なお、標準物質由来のシグナルを拘束条件に

設定する場合、標準物質由来シグナルが均等に広く分布するように標準物質を選択することで、探索空間削減効果を最も発揮することができる。

なお、探索空間を限定して探索効率を上げるためのアルゴリズムに対する改変方法としては、図5に示した探索空間の左上の出発点から右下の到達点に至る対角線の前後に、所定の幅を持った空間に限定するといった拘束条件も考え得る。しかしながら、この場合、どの程度限定すれば良いのかといった確実な前提知識

は一般に得られない虞がある。さらに、この場合、出発点や到達点が比較すべき多次元データ毎に大きくずれた場合には、求めるべき最適経路が限定された空間からはみ出す虞がある。例えば、クロマトグラフィーの溶出開始時刻は、大きく変動しうるので、この時刻を確実に観測できなければ、上記対角線の前後に所定の幅を持った空間に探索を限定する方法は適切とは言いがたい。

それに比べて、標準物質由来のシグナルを拘束条件に設定するようにアルゴリズムを改変する方法では、検体内部に添加した標準物質の出現した時点を揃えており、標準物質に由来するシグナルは最も確証のある参照点となる。その上、探索空間は最大 $\frac{1}{n+1}$ 近くまで減らせるため、確実性及び効率性の両面から見ても優

10 れているといえる。

さらに、標準物質由来シグナルで限定される 1 つないしいくつかの部分空間に限って最適経路探索を行うことで、部分的な最適プロファイル重ね合わせを求めることができる。この際に、前述の評価関数の値を、プロファイル重ね合わせの程度の指標として、プロファイル同士の類似性（あるいは距離）を測ることができる。多くの場合、主要なシグナルは限られた時間領域に集中的に現れるため、部分空間に限って最適経路探索を行って評価関数の値を求めて、プロファイル同士、さらにはそれらのプロファイルを生じさせた試料同士の類似性（あるいは距離）を効率よく求めることができる。

数多くの試料由来のプロファイルを重ね合わせて平均プロファイルを求めなければいけない場合や、あらかじめ試料の属性に関する情報が十分に得られていない場合は、まず部分空間に限って最適経路探索を行って試料間の類似性（または距離）を求めておき、その順番で逐次プロファイル重ねあわせを行い、あるいは試料同士のグルーピングを行うことが可能である。殊に、プロファイルの重ね合わせ処理を行う場合は、重ね合わせる順番に依存して結果が変化しうるので、できるだけ近いものから重ね合わせることが望ましいが、そのような処理を行う際にこの手法は有効である。

5. データ解析後の処理

2つの多次元データについて、最適な重ね合わせが得られたならば、補正されたパラメータについて、新たな補正後の値を生成する。特に、クロマトグラムの保持時間を補正した場合は、補正後の保持時間を探る。補正後の保持時間を探る方法としては、重ね合わせる2つの多次元データのうち一方を参照用データとし、他方の多次元データにおける保持時間を当該参照用データにおける保持時間に一致させる方法（非対称型）と、重ね合わせる2つの多次元データの両方を補正する方法（対称型）とを挙げることができる。特に対称型で補正後の保持時間を求めることが好ましい。

補正後の保持時間を求めて非対称型で2つの多次元データを重ね合わせる場合には、参照用データの保持時間軸にあわせるべく、一致した点に関しては参照用データの保持時間をそのまま用い、参照データの対応する点が得られなかつた場合は、その前後で一致した点を用いて内挿によって補正後の保持時間を決めることができる。

しかしながら、非対称型で2つの多次元データを重ね合わせるには、いずれの多次元データを参照用とすべきかが予め決まっている必要がある。例えば、標準物質のみを含むブランクを参照用データとすることも考えられるが、この場合は最初に重ね合わせに用いられた多次元データのプロファイルの影響がかなり大きくなる可能性が高い。

これに対して、補正後の保持時間を求めて対称型で2つの多次元データを重ね合わせる場合には、2つの多次元データ間で対応する点が得られたところについては、それぞれの保持時間の算術平均とする。2つの多次元データのいずれか一方だけのデータ点しか得られなかつた場合は、その前後の最も近い対応点の補正後の保持時間の組から、内挿によって補正後の保持時間を得る。内挿による補正が不可能な場合は、最も近くにある対応した点の補正後の保持時間をベースにして、データセット全体での平均時間スケールを係数とした外挿によって補正後の保持時間を得ることが可能である。

なお、この場合、前述のようにあらかじめ全ての多次元データ間の類似性（距離）を事前に計算した上で、近いものから順次重ね合わせても良い。

6. 出力処理

上記「5. データ解析後の処理」で得られたプロファイルの出力方法としては、以下の（1）及び（2）を挙げることができる。

（1）対応する点が得られなかつた箇所も含めて、全て出力する。

5 （2）対応する点に限って出力する。

いずれの出力方法を選択するかは、本発明に係る試料解析方法の使用目的に応じて適宜選択することができる。例えば、測定誤差を相殺するため同一試料を複数回に亘って測定した結果から平均を求める目的とする場合、又はきわめて類似した条件での複数試料の測定から代表的なプロファイルを得たい場合には上記（2）の出力方法が有効である。上記（2）の出力方法によれば、出力されるプロファイルは共通部分に限られるため、データの大きさを軽減して処理効率を上げることができる。

また、例えば異なった試料群の群間差を検出する場合には、上記（1）の出力方法を使用しなくてはならない。上記（1）の出力方法では、データの大きさは一般に増大するが、情報の損失は起こらない。

また、上記（1）の出力方法を選択する場合、さらに共通するプロファイルに重きを置いた重ね合わせを行うこともできる。この場合、以前の重ね合わせプロセスで対応する点に対して新たな点を対応させる場合は、評価関数のスコアを良くするような新たな項を当該評価関数に設けて、できるだけ同じ点に重なるように調整することも可能である。すなわち、例えば、上記式（I）で与えられる評価関数の最後に $-\mu \cdot \delta_m(i, j)$ のような新たな項を付与した新たな評価関数を用いて評価得点を算出する。そして、このような評価関数を用いる場合、以前の重ね合わせ点に対して対応が取れた場合には $\delta_m(i, j)$ を1、それ以外は0とする。

一方、本発明に係る試料解析プログラムの出力は次のような形式となっている。

- 25 • 重ね合わせ処理で新たに得られた点の情報
- 対応する入力データセット1（一方の多次元データ）の点の情報
- 対応する入力データセット2（他方の多次元データ）の点の情報

出力は、これらの情報が重ね合わせ処理の結果として得られたデータ点の数だけ繰り返された形をとる。但し、対応する点がない場合には、それぞれ入力デー

タセット 1 あるいは 2 の情報は存在しない。このように、出力には、対応する入力データセットの点の情報も含まれることにより、後述する例のように、最終的に得られた、重ね合わせプロファイルの各点が、元の多次元データのどれに由来するかを求めることが可能となる。なお、以上の情報の他にも必要であれば付加的情報をつけて出力することも可能である。

5 このようにして得られた重ね合わせ後のプロファイルについて、必要に応じてさらにいくつかのパラメータに関して集約あるいは量子化処理を行ってもよい。例えば、特に上記（1）のように全ての点を出力した場合は、時間軸分解能が求めるレベル以上に詳細になりすぎることもある。この場合は時間軸上きわめて近接した点同士はさらに集約して 1 つにした方が、後の処理上得策である。集約された点の強度は、集約前の個々の点の強度の加算で置き換えることができる。同様にして、 m/z 軸においても必要分解能以上に近接した点は集約することができる。但しこの操作は重ね合わせ処理を行う都度実行してもよいし、必要な重ね合わせをまず行ってから、最後に 1 回だけ行ってもよい。

15

7. イオン化強度の規格化と標準物質

本発明に係る試料解析方法においては、測定したイオン強度について、上記「3. データ解析」に先立って規格化を行うことが好ましい。以下にイオン強度の規格化について説明するが、イオン強度の規格化手法は何ら限定されるものではない。

20 具体的には、先ず、LC-MS 分析に結果として得られた RAW ファイルを、例えば Xcalibur™ のユーティリティソフトウェアを用いてテキストファイルに変換する。次に、C 言語および Perl 言語で作成されたプログラムにより、以下の一連のデータ処理を適用する。

(1) ノイズレベルのデータを除去するため、イオン強度が所定の値（例えば、 10^2 以下）以下のシグナルを除去する。

25 (2) 必要ならば、処理時間の節約のためにデータ点を集約する。具体的に、例えば、 m/z は 1 刻みに、保持時間は 0.2 刻みになるよう、元データの m/z 値および保持時間の値を丸め、同じ値を持つ (m/z 、保持時間) のデータ点は加算集計する。

(3) あらかじめ調べた m/z 値と保持時間から標準物質由来のシグナルを同定し、そのイオン強度値をもって測定値全体を除算することで規格化する。この際、1つないし複数の標準物質由来の複数のシグナルの平均値などの代表値を標準物質イオン強度値として用いる方法、予備実験などでシグナルの安定性を事前に検討したうえでもっとも安定なシグナルの値を用いる方法、などがある。

より具体的に、例えばニワトリ卵白リゾチームを標準物質とした場合、 m/z 値 715 近傍及び 877 近傍のシグナルを標準シグナルとすることができる。サンプルの測定データに対しては、 m/z については前後±1 の範囲で、保持時間に関しては m/z 715 (715 ± 1) のシグナルについては早期溶出シグナル群より、 m/z 877 (877 ± 1) のシグナルについては後期溶出シグナル群より探索することで、標準物質由来シグナルを探索することができる。なお、得られた値に 10^7 を乗じることで、標準物質由来のシグナル強度を 10^7 に補正するといった更なる補正を行ってもよい。

また、ペプチドそのもの、例えばペプチド T (Ala-Ser-Thr-Thr-Asn-Tyr-Thr) および β カソモルフィン 7 (Tyr-Pro-Phe-Pro-Gly-Pro-Ile) を標準物質として用いた場合、それぞれ m/z 値 859 近傍及び 791 近傍のシグナルを標準シグナルとすることはできる。前者のペプチドは比較的親水性であり、後者は疎水性である。本解析方法で保持時間による分離のために用いられている逆相クロマトグラフィーでは、前者の保持時間の値は低く、後者のそれはより高い。大部分のサンプル由来のペプチドの保持時間はこの 2 種類のペプチドの保持時間の間にある。サンプルの測定データに対しては、 m/z については前後±1 の範囲で、保持時間に関してはあらかじめ標準物質のみを測定して得られたクロマトグラムからおおよその値を見積もってから、その前後一定の範囲で探索することで、標準物質由来シグナルを見出すことができる。

本解析方法において標準物質としてペプチドを用いる場合、当該ペプチド以外の物質（例えば、夾雑物）を含まないことがシグナルノイズの検出を最小限に抑えるために重要である。したがって、天然物から抽出・精製して得られるものよりも、化学的に合成されたペプチド分子の方が、望ましい。また、ペプチド分子の性質については、あらかじめ設定された測定条件下で構造が安定なことや不溶

化しないことが重要である。ペプチド分子を構成するアミノ酸残基については、メチオニン、トリプトファン、ヒスチジンなど容易に酸化されうるアミノ酸残基を含まないこと、および塩基性の官能基を2つ以上含まないことが望ましい。とくに後者については、測定手段のMSで原理的に多価イオンを生じるエレクトロスプレーアイオン化法を用いた場合、ひとつの標準物質から価数の異なる複数のイオンシグナルが検出されることを避けるためである。

標準物質として蛋白質から加水分解あるいは化学的切断によって生じるペプチド断片を用いる場合、当該蛋白質に由来するペプチドイオンシグナルのうちで標準物質として採用したもの以外のペプチドイオンシグナルの強度は出来る限り低いことが望ましい。

以上(1)～(3)によって、測定されたイオン強度値を規格化することができ、複数の試料間におけるイオン強度の量的な比較を行うことができる。なお、測定されたイオン強度値の規格化は、上述した保持時間の補正に先立って行われるべきである。

15

8. 試料間の比較分析

本発明に係る試料解析方法によって m/z 、規格化されたイオン強度及び補正した保持時間からなる3次元データを用いて、試料中に含まれるタンパク質群等の各種成分分析をコンピュータ上で行うことができる。具体的に、成分分析としては、(a) 加算の方法、(b) 減算の方法を挙げることができる。

a. 加算の方法

上述したように、本発明に係る試料解析方法によって取得された複数の3次元データにおいては、保持時間のパラメータをそれぞれ適切に補正しているため、データ点間の対応関係を正確に取ることができる。したがって、複数の3次元データにおいて、データ点同士の規格化されたイオン強度値同士を足し合わせることができる。

b. 減算の方法

「a. 加算の方法」と同様に、本発明に係る試料解析方法によって取得された複数の3次元データにおいてはデータ点間の対応関係を正確に取ることができるた

め、データ点同士の規格化されたイオン強度値の差を求めることができる。

このように、本発明に係る試料解析方法によって取得された複数の3次元データについて、加算或いは減算することができるため、以下のような、成分分析のアプローチをコンピュータ上で実現することができる。

5 (1) 実験データを集計する際への応用：

1 つのサンプル由来の試料を、測定の便宜上、複数の分画に分割して測定した場合であっても、当該複数の分画それぞれから取得された3次元データにおいては、データ点間の対応関係を正確に取ることができる。したがって、上述した加算の方法に従って、全ての3次元データを足し合わせることができる。これにより元のサンプル全体に含まれる成分の解析等を行うことが可能となる。

なお、このような集計を行う際には、複数の分画全てに亘って足し合わせ、全体で1つのプロファイルにまとめる方法のほかに、あらかじめ指示された個数により、近接する数個のプロファイル毎に集約する方法もある。この場合、全体が n 個に分画され、そのうち近接する m 個ずつを集約した場合は $n-m+1$ 個のプロファイルが得られることになるが、その場合は得られたプロファイルについては、対応する分画同士以外はそれぞれ別個に扱って以下の操作を行うことになる。いずれにせよ、多次元分画法で複数の分画を得た場合も、そのおのおのが完全に重なりなくなることはまれで、多くの場合複数分画に亘るキャリーオーバーがあるので、このような集計操作は必要である。

20 (2) 複数のサンプルの測定結果の代表値を求める際への応用：

異なるサンプルに由来する複数の試料について測定した場合であっても、本発明に係る解析方法によれば、取得された複数の3次元データ間において、データ点間の対応関係を正確に取ることができます。したがって、上述した加算の方法に従って、全ての3次元データを足し合わせることができます。そして、得られた3次元データの総和をサンプル数で除算することで相加平均を求めることができます。なお、必要に応じて、各サンプルに重みを設定し、当該重みを反映させた重みつき平均を算出することもできる。

これによれば、例えば、同じ範暗に属すると考えられる複数のサンプルについて、当該範暗の代表値を求めることができます。

(3) 2つのサンプル間での測定結果の差分を求める際への応用：

例えば、同一のサンプル由来であるが状態の異なるときに採取した試料について測定した場合であっても、取得された2つの3次元データ間において、データ点間の対応関係を正確に取ることができる。したがって、上述した減算の方法に従って、2つの3次元データ間の差分を求めることができる。これにより、状態の変化に起因する試料中の成分変化を解析することができる。

また、例えば、上述した(1)に準じて、複数のサンプルを含む2群についてそれぞれ相加平均等の代表値を求めた上で、これら2群の代表値の差を求めることができる。得られた差については、統計的な検定などによって有意性を検討すること10で、各群に特異的な成分を同定することができる。

以上の(1)～(3)に示す成分分析アプローチは、本発明に係る試料解析方法によって得られた複数の3次元スペクトルデータを格納したデータベースを用いても良いし、当該データベースに格納されたデータと現実に得られたデータとを用いて行っても良い。いずれの場合であっても、上述した(1)～(3)に示す成分分析アプローチは、コンピュータを用いて容易に実現することができる。

このようにして、本発明に係る試料解析方法によって得られた、例えば群特異的シグナル成分については、得られたシグナル領域に範囲を限定したタンデムMS分析などにより、当該シグナルが由来する蛋白質群を同定することができる。すなわち、本発明に係る試料解析方法においては、試料をLC-MSにより分析したときに、特定のm/z値を持つペプチド分子イオンが検出された場合、当該イオンのCIDスペクトルを測定することができる。

そして、得られたCIDスペクトルをコンピュータに入力し、データベース検索ソフトウェアを用いて蛋白質一次構造データベース、ゲノム配列データベースやcDNA配列データベースから得られる蛋白質配列に対して検索する。このデータベース検索によって有意なヒットスコアを示した場合には、データベースに登録されたタンパク質或いはアミノ酸配列等の情報を得ることができ、得られたCIDスペクトルに対して当該情報を関連付けることができる。

例えば、上述した(3)の成分分析アプローチにおいて、各群に特異的な成分として同定されたシグナルに関するCIDスペクトルを測定することで、当該シグナ

ルが示すタンパク質群を同定することができる。

[実施例]

以下、実施例を用いて本発明をより詳細に説明するが、本発明の技術的範囲は
5 以下の実施例に限定されるものではない。

[実施例 1]

実施例 1 では、アミノ酸配列が既に知られているタンパク質のプロテアーゼ消化物を混合して得られるペプチド試料を LC-MS によって測定し、この測定によって得られた保持時間、 m/z 値およびイオン強度からなる三次元プロファイルに対して本発明に係るアルゴリズムを適用し、測定されたペプチド試料を定量的に特性づけた。また、実施例 1 では、比較定量のためのモデル実験としてアミノ酸配列が既に知られているタンパク質のプロテアーゼ消化物を混合したペプチド試料数種を各々 LC-MS によって測定し、本発明の試料解析方法を適用して各三次元プロファイルを比較することによって、各ペプチド試料に含まれるタンパク質の種類の違いが検出されることを示した。
15

ペプチド試料の調製

以下に列挙する 24 種類のタンパク質のトリプシン消化物を、本実施例におけるペプチド試料として調製した。(1) ウシキモトリプシノーゲン、(2) ウシカタラーゼ、(3) ウシカルボニックアンヒドラーーゼ、(4) ウシアポトランスフェリン、
20 (5) ウシカルボキシペプチダーゼ A、(6) ウシ血清アルブミン、(7) ウマシトクロム c、(8) ブタガンマ免疫グロブリン、(9) ウシヘモグロビン、(10) ウマミオグロビン、(11) ウシベータラクトグロブリン、(12) ウシデオキシリボヌクレアーゼ、(13) ウサギグリセルアルデヒド 3 リン酸デヒドログナーゼ、(14) トリコンアルブミン、(15) セイヨウワサビペルオキシダーゼ、(16) 枯草菌アルファアミラーゼ、(17) ウマグルタチオン S-トランスフェラーーゼ、(18) ウシグルタミン酸デヒドログナーゼ、一。(19) ウシラクトペルオキシダーゼ、(20) コウジカビアミログルコシダーゼ、(21) ウサギホスホリラーゼ B、(22) ウシベータガラクトシダーゼ、(23) ウサギ乳酸デヒドログナーゼ、(24) ニワトリ卵白リゾチーム。これらの消化物は Michrom BioResources 社より購入した。

これら 24 種類の各タンパク質のトリプシン消化物を以下に示すように混合し、合計 3 種類 (A 群～C 群) のペプチド試料を用意した。

A 群 : (1)、(2)、(7) ～ (24) の 20 種類のタンパク質のトリプシン消化物。A 群を特徴付けるタンパク質は (1) 及び (2) である。B 群 : (3)、(4)、(7) ～ (24) の 20 種類のタンパク質のトリプシン消化物。B 群を特徴付けるタンパク質は (3) 及び (4) である。C 群 : (5) ～ (24) の 20 種類の蛋白質のトリプシン消化物。C 群を特徴付ける蛋白質は (5) 及び (6) である。各群のサンプルは 3 つずつ調製した。

LC-MS 解析

各ペプチド試料の三次元プロファイルを得るために、以下に示す装置と操作によってペプチド試料を分析した (Kawakami, T. et al, Jpn. J. Electrophoresis 44: 185-190 (2000))。まず、減圧濃縮したペプチド試料を、トリフルオロ酢酸、アセトニトリルおよび水の混合比が 0.1:2:98 の溶媒 45 μ l に溶解した。これを溶解液とする。

次に、CTC Analytics 社製のオートサンプラー PAL LC-1TM を用い、Michrom BioResources 社製の MAGIC MSTM C18 キャピラリーカラム (内径 0.2mm、長さ 50mm、粒径 5 μ m、孔径 200 オングストローム) に溶解液 20 μ l を導入した。ペプチドの溶出は MAGIC 2002TM HPLC システム (Michrom BioResources 社) を用いて行った。このときの HPLC 移動相 A は蟻酸、アセトニトリル及び水を 0.1:2:98 の容積比で混合した溶媒であり、対して移動相 B の混合比は 0.1:90:10 であった。そして移動相 B の濃度を 5%から 85%まで直線勾配で上げ、ペプチド断片を連続的に溶出した。このときの流速は約 1 μ l/min とした。LC の溶出液は、New Objective 社製の PicoChipTM ニードル (内径 20 μ m) を介し、LCQTM イオントラップ型質量分析計 (ThermoQuest 社) のイオン源に直接導入した。NanoESI ニードルの位置は加熱キャピラリーとの距離を微調整できるようになっている。また、スプレー電圧はニードルではなく、溶離液に直接荷電するようにした。噴霧のためにガスは使用せず、スプレー電流は 3.0mA とした。これを各群 3 回ずつ行うことによって、各試料に対応する三次元パラメータの集合体、3 群計 9 通りを得た。このデータセットをそれぞれ A1、A2、A3 (A 群)、B1、B2、B3 (B 群)、C1、C2、C3 (C 群) とした。

三次元パラメータ集合体を含むファイルは Xcalibur™のユーティリティソフトウェアを用いてテキストファイルに変換した。C 言語及び Perl 言語で作成されたプログラムにより、以下の①～⑤のデータ処理を実行した。

①ノイズレベルのデータを除去するため、イオン強度が 10^2 以下のシグナルを除去した。
5

②処理時間の節約のためにデータ点を集約した。具体的には、 m/z は 1 刻みに、保持時間は 0.2 刻みになるよう、元データの m/z 値および保持時間の値を丸め、同じ値を持つ m/z と保持時間の 2 つ組みで指定されるデータ点は加算集計した。

③標準物質であるニワトリ卵白リゾチーム由来のシグナルを同定した。すなわち、
10 予備実験で実測した標準物質の m/z 値および保持時間の値の前後ある範囲内で、最も高いイオン強度を与えるデータ点を探し、次いでそのデータ点を中心に、イオン強度値が単調減少しかつ 0 より大きな範囲にあるデータ点を拾い、これらを標準物質由来のシグナルによるデータ点であると見なした。標準物質由来のシグナルの総イオン強度値としては、標準物質由来シグナルと見なされたデータ点のイオン強度の総和をもって当てた。具体的には、ニワトリ卵白リゾチーム由來の
15 m/z 値が 715 近傍及び 877 近傍のシグナルを標準シグナルとし、サンプルの測定データからこれらの標準物質由来シグナルを探索する際には、 m/z については前後 ±1 の範囲で、保持時間に関しては m/z 715 のシグナルについては 6～16 分の範囲で、 m/z 877 のシグナルについては 13～23 分の範囲で探索した。

20 ④得られた標準物質由来シグナルの総イオン強度値をもって、各シグナルのイオン強度を除算し、得られた値に 10^7 を乗じることで、標準物質由来のシグナル強度を 10^7 に補正した。

⑤便宜上、 m/z 715 のシグナルと m/z 877 のシグナルのピーク位置がそれぞれ保持時間に関して 10 分、20 分となるよう、保持時間軸を線型変換した。

25 次に、A、B 及び C 群それぞれ 3 例ずつあるサンプルから得られた三次元プロファイルの代表点を求めた。すなわち上述したように、同じ群に属するサンプルを集め、 m/z および保持時間の重なる点のイオン強度は加算して集計した。

また本例で使用したスコアは高いほどよいとする得点として計算し、算出式の係数は次の通り。イオン強度差については、それぞれの常用対数の差の絶対値に

対して係数-1 をかけたものを使用した。保持時間の差については、差の絶対値に対して係数-1000 をかけたものを使用した。また各群間で対応するデータ点のシグナルがともに標準物質由来であった場合の加算点は 50000 点とした。一方の群において対応する保持時間の点がなかった場合の失点は 5000 点とした。本実施例ではこれらを単純に加算してスコアとした。

次いで、上述したように、A-B 群間、B-C 群間及び C-A 群間で差を求めた。得られた差の有意性は t 検定によって有意水準 0.1% の両側検定で検討した。

その結果、保持時間を補正した三次元データを比較することによって、A、B 及び C の各群から以下に列挙する m/z 値を持つペプチド分子イオンが各群に特異的なシグナルとして検出された。

A 群 : 495, 524, 546, 560, 671, 696, 779, 845, 871, 908, 962 等。

B 群 : 451, 464, 509, 513, 546, 555, 583, 585, 626, 635, 649, 653, 701, 720, 723, 740, 741, 753, 768, 789, 819, 821, 847, 873, 886, 922, 928, 952, 966, 973, 978, 1057, 1230 等。

C 群 : 636, 670, 674, 679, 683, 718, 734, 735, 770, 824, 870, 918 等。

また、本実施例では、特異的なシグナルとして検出されたペプチド分子イオンの CID スペクトルを得るために、各試料を LC-MS/MS 分析にかけた。分析条件は以下に示す操作以外は、上述した通りとした。すなわち、LC-MS/MS 分析に際しては、イオントラップ型質量分析計の測定条件を変更し、上に列挙した m/z 値を持つペプチド分子イオンが検出された場合に当該イオンの CID を必ず行うように測定条件を設定して試料の測定を行った。

その結果、各ペプチド分子イオンから得られた CID スペクトルを、Matrix Science 社のデータベース検索ソフトウェアである MASCOTTMを用いて、SWISS-PROT 蛋白質配列データベースに対して検索したところ、各群において特異的なタンパク質由来ペプチドとして添加した各群 2 種類（すなわち、A 群においては上記 (1) 及び (2) のタンパク質、B 群においては (3) 及び (4) のタンパク質、C 群においては (5) 及び (6) のタンパク質）、合計 6 種類はいずれも有意なヒットスコアをもって同定された。このことから、本実施例で行った試料解析方法の妥当性が示された。

[実施例 2]

実施例 2 では、所定の濃度組成の蛋白質混合物中に、濃度を変化させた別の蛋白質標品を混合して得られた試料をプロテアーゼ消化し、LC-MS によって測定することで得られた保持時間、 m/z 値、イオン強度からなる 3 次元データに対して本発明に係る方法を適用し、異なる濃度の試料の測定で得られた 3 次元データ同士を比較することで、定量的に変動するシグナルを検出した。これによって、本手法で定量的に変化する物質を検出できることを示したものである。

試料とその調製

以下に列挙する 6 種類のタンパク質のトリプシン消化物を、本実施例におけるペプチド試料として調製した。(1) ウシカタラーゼ、(2) ウシベータラクトグロブリン、(3) ウシラクトペルオキシダーゼ、(4) ウマグルタチオン S-トランスフェラーゼ、(5) セイヨウワサビペルオキシダーゼ、(6) ウシ血清アルブミン。これらのタンパク質はシグマ社より購入した。

これらのタンパク質は水溶液中でブタトリプシン（プロメガ社より購入）と反応させることにより、トリプシン消化物とした。

これら 6 種類の各タンパク質のトリプシン消化物を以下に示すように混合し、合計 7 種類のペプチド試料を用意した。

①：(1)～(5) はそれぞれが一測定当り 500 フェムトモル、(6) は 0 フェムトモル。

②：(1)～(5) はそれぞれが一測定当り 500 フェムトモル、(6) は 10 フェムトモル。

③：(1)～(5) はそれぞれが一測定当り 500 フェムトモル、(6) は 50 フェムトモル。

④：(1)～(5) はそれぞれが一測定当り 500 フェムトモル、(6) は 100 フェムトモル。

⑤：(1)～(5) はそれぞれが一測定当り 500 フェムトモル、(6) は 500 フェムトモル。

⑥：(1)～(5) はそれぞれが一測定当り 500 フェムトモル、(6) は 1 ピコモル。

⑦：(1)～(5) はそれぞれが一測定当り 500 フェムトモル、(6) は 5 ピコモル。

各群のサンプルは 5 測定分ずつ調整した。上記各試料に対し、さらにペプチド T および β カソモルフィン 7 をそれぞれ 10 ピコモルおよび 1 ピコモルずつ標準物質として混入した。

LC-MS 解析

5 各ペプチド試料の三次元データを得るために、以下に示す装置と操作によってペプチド試料を分析した (Kawakami, T. et al, Jpn. J. Electrophoresis 44: 185-190 (2000))。まず、減圧濃縮したペプチド試料を、トリフルオロ酢酸、アセトニトリル及び水の混合比が 0.1:2:98 の溶媒 45 μ l に溶解した。これを溶解液とする。

10 次に、CTC Analytics 社製のオートサンプラー PAL LC-1TMを用い、Michrom BioResources 社製の MAGIC MSTM C18 キャピラリーカラム(内径 0.2mm、長さ 50mm、粒径 5 μ m、孔径 200 オングストローム)に溶解液 20 μ l を導入した。ペプチドの溶出は MAGIC 2002TM HPLC システム (Michrom BioResources 社) を用いて行った。このときの HPLC 移動相 A は蟻酸、アセトニトリル及び水を 0.1:2:98 の容積比で混合した溶媒であり、対して移動相 B の混合比は 0.1:90:10 であった。そして移動相 B の濃度を 5%から 85%まで直線勾配で上げ、ペプチド断片を連続的に溶出した。このときの流速は約 1 μ l/min とした。LC の溶出液は、New Objective 社製の PicoChipTMニードル (内径 20 μ m) を介し、LCQTMイオントラップ型質量分析計 (ThermoQuest 社) のイオン源に直接導入した。NanoESI ニードルの位置は加熱キャピラリーとの距離を微調整できるようになっている。スプレー電圧はニードルではなく、溶離液に直接荷電するようにした。噴霧のためにガスは使用せず、スプレー電流は 3.0mA とした。また、質量分析計でのスキャン回数を細かく取るために Turbo Scan 方法を適用した。この測定を各群 5 回ずつ行うことによって、各試料に対応する三次元パラメータの集合体、7 群計 35 通りを得た。図 1 に得られたプロファイルの例を示す。

データ処理

三次元データを含むファイルは XcaliburTMのユーティリティソフトウェアを用いてテキストファイルに変換した。C、C++言語及び Perl 言語で作成されたプログラムにより、以下の①～⑦のデータ処理を実行した。

①ノイズレベルのデータを除去するため、イオン強度が 10^2 以下のシグナルを除去した。

②標準物質であるペプチド T および β カソモルフィン 7 由来のシグナルを同定した。すなわち、予備実験で実測した標準物質の m/z 値および保持時間の値の前後 5 ある範囲内で、最も高いイオン強度を与えるデータ点を探し、その周囲で規定の範囲内にあるシグナルについて、さらにその点を頂点とするガウス分布で近似される強度範囲内に収まるシグナルを拾い上げ、それらを全て標準物質由来シグナルとみなした。より具体的には、 m/z は 858.9 および 791.0 を中心に前後 ± 2 を限度とし、また保持時間は 9 および 25 分を中心に前後 ± 6 を限度として、上述 10 の手順により探索した。従って使用した標準物質由来シグナルは m/z 858.9, 保持時間 9 分、および m/z 791.0, 保持時間 25 分近傍に存在する 2 つである。なお、イオン強度補正を行う際には標準物質由来シグナル強度を全て足し合わせ、その値が 10^9 となるように規格化した。また時間軸補正の際の動的計画法探索空間上の拘束点としては、上述 2 つの標準物質シグナルのうち、それぞれ強度ピーク 15 を与える 1 点ずつの計 2 点を選定した。

③BSA の濃度を変えた 7 種類の試料それぞれにつき、5 回の測定結果得られたプロファイルを平均化するため、本発明によるプロファイル重ね合わせプログラムを用い、7 群の試料それぞれについての代表プロファイルを求めた。重ね合わせの際の評価関数のパラメータは以下のとおり：

20 上記 (I)において、時間軸上の差（絶対値）のペナルティ $\alpha = 1.0$ 、シグナル強度の差のペナルティ $\beta = 0.1$ （但しシグナル強度さは常用対数に変換後の差の絶対値とした）、点の一致に対するボーナス点 $\sigma = 0$ 、不一致点に対するペナルティ $\pi = 100$ 、標準物質由来シグナルに対する一致のボーナス点 $\theta(i, j) = S_m = 1000$ 。また重ね合わせプロファイルの出力オプションは、一致点のみとした。

25 図 8 に、BSA 濃度が 500 フェムトモルの試料⑤の 5 回の測定結果から得られたプロファイルのうち、 m/z 620.0 近傍、残留時間 15 から 19 分の間のクロマトグラムを示す。時間軸上少しづつずれた灰色の 5 本の波形が、重ね合わせ処理前の 5 回の測定結果であり、実線がこれら 5 つに重ね合わせ処理を施してすべてのシグナルを足し合せた結果である。図に示したように時間軸上の揺らぎが補正さ

れて 1 つの大きなピークとして扱えるようになった。

④次に、群間で有意に変動しているシグナルを拾い出すため、前述のプロセスで得た 7 群の代表プロファイルを、さらに重ね合わせ処理によってアライメントをとり、最終的に 1 つの集約プロファイルを得た。このときの評価関数のパラメータは以下のとおり：

上記式(I)において、時間軸上の差（絶対値）のペナルティ $\alpha = 1.0$ 、シグナル強度の差のペナルティ $\beta = 0.1$ （但しシグナル強度を常用対数に変換後の差の絶対値とした）、点の一致に対するボーナス点 $\sigma = 0$ 、不一致点に対するペナルティ $\pi = 100$ 、標準物質由来シグナルに対する一致のボーナス点 $\theta(i, j) = S_m = 1000$ 。また重ね合わせプロファイルの出力オプションは、不一致を含む全ての点とした。

⑤試料間の相互比較に必要十分な分解能を保ちつつデータ点を集約するため、 m/z 範囲 ± 0.75（絶対範囲 ± 2）、保持時間範囲 ± 1.25（絶対範囲 ± 4）の範囲内にある点について、次の条件を満たしている点については 1 つに集約した。即ち、シグナル強度の高い順に前述範囲内の全てのデータ点をチェックし、これらがピークシグナルを頂点とするガウス分布で近似される範囲内に入ると判定されたものについてはそれらを集約した。

図 9 に、BSA 濃度が異なる 7 種類の試料ごとの、時間軸補正および集約後のプロファイルの例を示す。図では特定の m/z 値（この例では 752）で切った断面を、時間軸に沿って強度をプロットしたクロマトグラムとして示している。17 分および 19 分近傍の集約シグナルは、BSA 濃度が最も高い 06 の試料（図中、“DS:Sp1 06-Ave” として示す。BSA 5 ピコモル）が最も高いピークを示し、順次 Sp1 05、Sp1 04, … と続いているため、これらは BSA 由来のシグナルと判定できる。一方、25 分近傍の緩やかなピークはどの試料でも同様に出ていたため、BSA 以外の共通物質あるいはバックグラウンド由来と判定できる。

⑥前述の集約プロファイルを出発点として、順次⑤、④、③のプロセスを逆にたどることにより、最終的に得られた集約プロファイル上の点が、測定結果の 7 群 × 5 測定のいずれに由来するかをたどることができる。集約プロファイル上の各点について、測定結果データに戻ることで、群ごとのプロファイルや群間の差プロファイルを得ることができる。図 10 に BSA 濃度が 500 フェムトモルの試料⑤と、0

フェムトモルの試料①との差プロファイルを示す。m/z—保持時間平面より上に伸びている線は試料⑤で、逆に下に伸びている線は試料①で強く観測されたシグナルである。

⑦BSAを入れていない試料①から得られた集約プロファイルを基準とし（基準プロファイル）、残りの②から⑦までの BSA 濃度が異なる 6 種類の試料から得られた集約プロファイルそれぞれについて（対象プロファイル）、次のような条件を満たすデータ点を求めた。即ち、（1）対象プロファイル上の点は、5 回の測定すべてよりデータが得られていること。（2）集約した対象プロファイル上の点のシグナル強度は 10^6 以上であること。（3）対象プロファイル上の各点について、基準プロファイル上の対応する点の強度を引いた差が 0 以上であること。（4）対象プロファイルの各点について、基準プロファイル上の対応する点とのシグナル強度の差が 10000 未満であるという帰無仮説が、片側 t 検定結果によって有意水準 0.5% で棄却されること。

上記の条件で選別されたシグナルを、試料⑤の場合について図 11 に示す。ここでは 127 個のシグナルが前述条件に合致したものとして残った。プロットマークの大きさは、試料⑤のプロファイルにおけるシグナル強度を表す。このうち、プロットマーク○のものは、後述するプロセスで BSA シグナルと対応付けられたもの、×のものはそうでなかつたものを示している。

これら 127 個のシグナルについて、別途行った BSA 消化物の MS/MS 実験とつき合わせたところ、このうち 103 個が BSA 由来シグナルと一致した。即ち、検出された 127 個のシグナルのうちの少なくとも 81% が真に求めたいものであったといえる。

同様にして、他の BSA 濃度の試料から検出されたシグナルについて、前述の同じ条件で選別されたシグナルを BSA 由来シグナルと付き合わせたところ、対象プロファイル②から⑦について、それぞれ 65、64、75、81（既述）、76、48% のシグナルが BSA 由来シグナルと判定された。なお最後の⑦試料（BSA 濃度 5 ピコモル）の正答率が落ちているが、これは高濃度 BSA に由来する強度の強いシグナルが多数存在することによるプロファイルの閾値が変わったことで、偽陽性シグナルが増大したことに起因すると考えられる。実際、他の濃度の試料とほぼ同程度

の数のシグナルが選ばれるよう、選定条件の（2）を 3×10^6 に調整すると、正答率は75%になった。

なお、本発明に係る工夫の1つに、動的計画法の探索空間節減があるが、その効果をCPU時間の実測によって評価したところ、試料⑤の5つの測定結果を重ね合わせる場合を例とすると、CPU時間の比にして43～45%の削減効果が得られた。

ここでは2種類の標準物質由来シグナルを用いているので、もし完全に均等にシグナルが分布していれば1/3の時間短縮が期待できるが、実際多くのシグナルは2つの標準物質シグナルの間に存在するので、探索空間の分割は不均等である。このことを勘案すると、45%程度の削減はおよそ期待通りであり、実用上も十分効果があると考えられる。

以上により、本発明に係る試料解析方法が、試料中で量的に変化するシグナルを実用上有用な確度で検出できることを示し、また本発明における計算手法の工夫が有用であることを示した。

15 [実施例3]

実施例3では、実際の患者由来組織サンプルを用いて、いくつかの病態群間で有意に変動している蛋白質由来シグナルを求め、それを元にさらにMS/MS解析を行ってそのうちのいくつかの蛋白質を同定することにより、本方法の有効性、特にバイオマーカーの探索等にも有効であることを示したものである。

具体的には、肺の腺癌を対象として、外科的に切除された組織を用い、組織より後述する手法で蛋白質を抽出し、測定した。得られたプロファイルを、後日病理診断によってリンパ節転移ありと判定された群となしと判定された群にわけ、両群間で有意に変動しているシグナルを拾い出し、それらについてMS/MS解析を行って蛋白質を同定した。

25 試料

試料は、36名の異なった肺癌患者より、外科的に切除された肺の組織片を用いた。病理診断により、これらの患者は腫瘍径の大きな群と小さな群、および所属リンパ節への転移が認められる群と認められない群の合計4群に分けられる。

腫瘍径小でリンパ節転移なしの群には10例、腫瘍径大でリンパ節転移なしの群

には 11 例、腫瘍径小でリンパ節転移ありの群には 12 例、腫瘍径大でリンパ節転移ありの群には 3 例にぞれぞれ分類された。

試料調製および蛋白質分画

ドデシル硫酸ナトリウム(SDS)-ポリアクリルアミドゲル電気泳動(PAGE)用の試

5 料緩衝液中にて各組織片を破碎した。試料緩衝液の組成は以下のとおりである。

62.5mM トリス-塩酸(pH6.8)、2%w/v SDS、5%v/v 2-メルカプトエタノール、10%v/v

グリセリン、0.0025%w/v ブロモフェノールブルー。この懸濁液を室温にて 30 分

間振とうしたのち、上清と沈殿に遠心分離した。上清の蛋白質の濃度を Lowry 法

の変法によって測定した。蛋白質 100 μg 分の試料上清に対して同じ組成の試料緩

10 衡液を加え、総液量を 50 μL にした。この溶液に対して 1M トリス水溶液を加え、

pH を 8.8 にした。システイン残基の還元アルキル化のため、2 μL の 400mM ジチオ

トレイトールを加え、60°C にて 30 分間保温した。つぎに 10 μL の 400mM ヨードア

セトアミド溶液を加え、室温・暗条件下で 60 分間放置した。約 5 μL の 1.0N 塩酸

を加え、pH を 6.8 に戻した。この溶液を Laemmli の SDS-PAGE にかけた。このと

きに用いたポリアクリルアミドゲルは不連続の緩衝液系、すなわち上部の濃縮ゲ

ル(pH6.8)と下部の分離ゲル(pH8.8)から成了た。ポリアクリルアミドゲルの濃度

は各々 4% と 12.5% であり、全体の大きさは幅 14cm、高さ 14cm、および厚さ 1mm

であった。電気泳動時の電流は一定の 10mA であった。色素ブロモフェノールの泳

動フロントが濃縮ゲルと分離ゲルの界面から分離ゲルの 48mm まで達したところ

20 で泳動を停止した。ポリアクリルアミドゲルを 40% メタノール、10% 酢酸水溶液

中で振とうし、ポリアクリルアミドゲル内で分離されている蛋白質を固定した。

このあと、ポリアクリルアミドゲルを水で 2 回洗浄した。洗浄したポリアクリル

ミドゲルを試料当たり 24 個のゲル片に切り出し分画した。すなわち、泳動方向に

対して垂直の向きに 2mm の等幅で梯子状に切り出し、各切片をさらに一辺約 1mm

25 のサイコロ状に分割した。

標準蛋白質の調製

内部標準蛋白質はゲル中に固定された状態で各試料ゲル分画に加えた。最初に

12.5% のアクリルアミド、0.1% SDS、375mM トリス塩酸(pH8.8)よりなる水溶液に

対して卵白リゾチーム水溶液を混合し、さらに N,N,N',N'-テトラメチルエチレン

ジアミンと過硫酸アンモニウムを加えてガラス板にはさまれた 1mm 幅の間でアクリルアミドを重合した。この内部標準蛋白質を含むゲルを直径 1.5mm の円形に繰り抜いた。ゲル切片当たり 2.5pmol の卵白リゾチームが含まれるように、重合前の水溶液の蛋白質濃度をあらかじめ計算しておいた。

5 プロテアーゼ消化

上に示した一定量の標準蛋白質を含むゲルの繰り抜き切片を各試料ゲル分画に 1 個ずつ加えた。つぎに各分画ごとにゲル片を十分量の水で洗浄してからアセトニトリルで脱水した。ゲル片に残った水とアセトニトリルを減圧留去してからトリプシン水溶液をゲル片全部が浸かる程度に加え氷中で 45 分間放置した。ゲル中に染み込まなかった水溶液を除き、50mM 重炭酸アンモニウム水溶液をゲル片全部が浸かる程度に加え、37°C にて 16 時間保温し消化反応を行った。ゲル片に含まれるペプチド断片の抽出は 25mM 重炭酸アンモニウム/50%アセトニトリル水溶液で 1 回、続いて 5%蟻酸/50%アセトニトリル水溶液で 2 回行い、抽出溶液は 1 個の容器にまとめて減圧濃縮した。

15 LC-MS 解析

各ペプチド試料の三次元プロファイルを得るために、以下に示す装置と操作によってペプチド試料を分析した (Kawakami, T. et al, Jpn. J. Electrophoresis 44: 185-190 (2000))。まず、減圧濃縮したペプチド試料を、トリフルオロ酢酸、アセトニトリル及び水の混合比が 0.1:2:98 の溶媒 45 μl に溶解した。これを溶解液とする。

次に、CTC Analytics 社製のオートサンプラー PAL LC-1™ を用い、Michrom BioResources 社製の MAGIC MS™ C18 キャピラリーカラム (内径 0.2mm、長さ 50mm、粒径 5 μm、孔径 200 オングストローム) に溶解液 20 μl を導入した。ペプチドの溶出は MAGIC 2002™ HPLC システム (Michrom BioResources 社) を用いて行った。このときの HPLC 移動相 A は蟻酸、アセトニトリル及び水を 0.1:2:98 の容積比で混合した溶媒であり、対して移動相 B の混合比は 0.1:90:10 であった。そして移動相 B の濃度を 5%から 85%まで直線勾配で上げ、ペプチド断片を連続的に溶出した。このときの流速は約 1 μl/min とした。LC の溶出液は、New Objective 社製の PicoChip™ ニードル (内径 20 μm) を介し、LCQ™ イオントラップ型質量分析計

(ThermoQuest 社) のイオン源に直接導入した。NanoESI ニードルの位置は加熱キャピラリーとの距離を微調整できるようになっている。スプレー電圧はニードルではなく、溶離液に直接荷電するようにした。噴霧のためにガスは使用せず、スプレー電流は 3.0mA とした。

5 データ処理

得られた LC-MS プロファイルデータは、36 サンプル×24 バンド分の、合計 864 個である。これらのプロファイルデータについて、Xcalibur™のユーティリティソフトウェアを用いてテキストファイルに変換した後、C, C++ 言語及び Perl 言語で作成されたプログラムにより、以下のような手順で解析を行った。

10 ①ノイズレベルのデータを除去するため、イオン強度が 10^2 以下のシグナルを除去した。

②処理時間の節約のために、 m/z と保持時間を量子化することでデータ点を集約した。具体的には、保持時間が約 1 刻みになるよう、シグナル強度が大きなものから順次、時間差 1 を最大限度に近傍を探し、単調減少する範囲までを 1 つのシグナルとしてまとめられる範囲とした。また m/z は 1 刻みになるよう、元データの m/z 値の値を丸め、前述の時間範囲内で m/z が同じ値を持つデータ点は加算集計した。

20 ③標準物質であるニワトリ卵白リゾチーム由来のシグナルを同定した。すなわち、予備実験で実測した標準物質の m/z 値および保持時間の値の前後ある範囲内で、最も高いイオン強度を与えるデータ点を探し、次いでそのデータ点を中心に、イオン強度値が単調減少しかつ 0 より大きな範囲にあるデータ点を拾い、これらを標準物質由来のシグナルによるデータ点であると見なした。標準物質由来のシグナルの総イオン強度値としては、標準物質由来シグナルと見なされたデータ点のイオン強度の総和をもって当てた。具体的には、ニワトリ卵白リゾチーム由來の m/z 値が 715 近傍及び 877 近傍のシグナルを標準シグナルとし、サンプルの測定データからこれらの標準物質由来シグナルを探索する際には、 m/z については前後 ± 1 の範囲で、保持時間に関しては m/z 715 のシグナルについては 10 分 ± 5 分の範囲で、 m/z 877 のシグナルについては 18 分 ± 5 分の範囲で探索した。但し、得られた標準物質由来シグナルの絶対強度、全シグナル中での相対強度、および

2 種類の標準物質由来シグナルの強度比のいずれかが他のものに比べてかけ離れているものについては、個別にプロファイルのプロットを確認のうえ、標準物質由来シグナルと考えられるシグナル群のピークが探索時のパラメータの中心点に来るよう調整した上で取り直した。得られた標準物質由来シグナルの総イオン強度値をもって、各シグナルのイオン強度を除算し、得られた値に 10^7 を乗じることで、標準物質由来のシグナル強度を 10^7 に補正した。さらに、便宜上、 m/z 715 のシグナルと m/z 877 のシグナルのピーク位置がそれぞれ保持時間に関して 10 分、20 分となるよう、保持時間軸を線型変換した。

④ SDS-PAGE によって分画した 24 バンド分のプロファイルについて、バンド間にまたがって存在する蛋白質の定量性を保証するため、全バンドについて重ね合わせたプロファイルをもって、各試料のプロファイルとして扱った。具体的には、本発明に係る試料解析プログラムの、プロファイル重ね合わせ機能を用い、隣接バンド間のプロファイルを逐次重ね合わせて加算集計した。即ち、まずはバンド $1+2$, $2+3$, $3+4$, …, $23+24$ のように、共通バンドを有する形で隣接バンドを重ねあわせ、次段階では $1+2$ と $2+3$ から $1 \sim 3$ の重ね合わせ得るという形で、常に最低 1 バンドは重なるようにして重ねてゆくと、6 段階の重ね合わせ操作により、最終的には $1 \sim 17$ と $9 \sim 24$ の重ね合わせによって、全バンドの重ね合わせを得た。なお、重複して重ね合わせたバンドに関しては、最終段階で重複回数を除算することで定量性を保つよう調整した。

20 重ね合わせの際のパラメータは以下のとおり：

上記式 (I) において、時間軸上の差（絶対値）のペナルティ $\alpha = 1.0$ 、シグナル強度の差のペナルティ $\beta = 1.0$ （但しシグナル強度を常用対数に変換後の差の絶対値とした）、点の一致に対するボーナス点 $\sigma = 100$ 、不一致点に対するペナルティ $\pi = 10$ 、標準物質由来シグナルに対する一致のボーナス点 $\theta(i, j) = S_m = 1000$ 。また重ね合わせプロファイルの出力オプションは、不一致を含む全ての点とした。さらに重ね合わせの処理が終わるごとに、保持時間と m/z がそれぞれ分解能 1.0 および 1.0 となるようデータ点の集約操作を行った。

⑤ リンパ節転移のある群とない群を特徴付ける蛋白質を探すため、前述の試料の 4 種類の分類に従って、まずは群内でプロファイル同士を重ね合わせて集約プロ

ファイルを得、次いで群間について同様にプロファイル重ね合わせ操作を行った。このときの重ね合わせ処理の際のパラメータは、前述のバンド間重ね合わせ処理と同じとした。また重ね合わせ順序については、群内については事前に総当たりで行っておいた同一パラメータの下での重ね合わせ処理の評価関数スコアをもとに、
5 近いものから順次重ね合わせた。群間については、まずリンパ節転移のある群内の腫瘍径の違いの 2 群と、リンパ節転移のない群内での腫瘍径の違いの 2 群を重ねた後、最終的にリンパ節転移のある群とない群を重ねた。

図 1 2 に最終的な重ね合わせプロファイルについて、リンパ節転移陽性群に表されたシグナルを上方向に、陰性群に存在したシグナルを下方向にプロットしたもの¹⁰を示す。

⑥前述の重ね合わせた順序を逆にたどることで、重ね合わせの出発点となっている 36 検体 × 24 バンドの元データにたどれるようにし、最終的な全てを重ね合わせたプロファイル上の各点について、元データとの対応を取れるようにした。
⑦最終的な集約プロファイル上の各点について、リンパ節転移ありの検体由来の
15 データと、なしの検体由来のデータをそれぞれ集計し、これら両群の平均値の差について両側 t 検定を行い、両群間の平均値の差と検定による p 値を得た。

図 1 3 に、上記検定で p 値が 0.005 未満となった点を、図 1 2 と同様のプロットで示す。この段階で 5,889 個のシグナルが得られた。

⑧上述のようにして選別されたシグナルについて、このシグナルの情報をもとに
20 ターゲット MS/MS を行い、あるいは別途行った MS/MS 解析から、蛋白質同定ソフト MASCOT™ を用いて、シグナルが由来するところの蛋白質同定を行った。図 1 4 に、この同定によって蛋白質情報と関連付けることができたシグナルを示す。約半数の 2,753 個のシグナルについて、何らかの既知蛋白と関連付けることができた。

25 最後に、上述の既知蛋白質と関連付けられたものの中から、癌の転移に何らかのかかわりを持つといわれている蛋白質の一部をリストにして示す(図 1 5)。このように、癌転移に関わると考えられる蛋白質が有効に発見しうることを示せた。

以上より、本発明に係る試料解析システムおよびプログラムが、実際の臨床検体を用いた解析に有効であり、殊に病理的・臨床的な差異を蛋白質の量的な差異

に結び付けて検出することができ、またその結果を利用して有効に蛋白質同定に繋げられることから、バイオマーカーの探索や新規診断方法の開発にも有用であることを示したものである。

- 5 本明細書で引用した全ての刊行物、特許および特許出願をそのまま参考として本明細書に取り入れるものとする。

産業上の利用の可能性

以上、詳細に説明したように、本発明に係る試料解析方法及び試料解析プログ

- ラムによれば、試料に含まれる成分を分析するに際して、優れた分析能を達成することができる。したがって、本発明によれば、分析対象の試料中に含まれる多数の成分を網羅的に解析する場合に非常に有効且つ有益な試料分析方法及び試料分析プログラムを提供することができる。

特に本発明に係る試料解析方法及び試料解析プログラムによれば、実際の臨床

- 5 検体を用いて、何らかの疾患の病態の違いに関係する物質を探査する目的にたいへん有効であり、これによりバイオマーカーの探索や診断法の開発ができるという点でも有用性はきわめて大である。

請求の範囲

1. 試料の分析の結果として得られた多次元データにおける、少なくとも 1 次元のパラメータを補正する工程 a と、

上記工程 a により得られる補正後のデータを、複数の試料について比較する工程 b とを含む試料解析方法。

2. 上記多次元データは、クロマトグラフィー質量分析の結果として得られる、質量／電荷比を示すパラメータと、イオン強度を示すパラメータと、保持時間を見すパラメータとからなる 3 次元データであり、上記工程 a では保持時間を示すパラメータを補正することを特徴とする請求の範囲第 1 項記載の試料解析方法。

3. 上記工程 a では、補正対象のパラメータを除いたパラメータに関するプロファイルを参照プロファイルとし、複数の試料間で複数の参照プロファイルに関する配置類似性の尺度となる評価関数を用いて、当該評価関数の値を最適化する最適解探索問題として各プロファイルの配置を求めることが特徴とする請求の範囲第 1 項記載の試料解析方法。

4. 上記評価関数は、以下の①～⑤からなる群から選ばれる 1 以上の項で定義されることを特徴とする請求の範囲第 3 項記載の試料解析方法。

①補正対象のパラメータに関するプロファイル間の類似性及び/又は距離に関する項

②参照プロファイルに関するプロファイル間の類似性及び/又は距離に関する項

③比較対象となるプロファイル間でのデータ点の一致の程度に関する項

④比較対象となるプロファイル間でのデータ点の不一致の程度に関する項

⑤比較対象となるプロファイル間での標準物質由来シグナルの一致又は不一致の程度に関する項

⑥比較操作を繰り返す際に、過去の比較における一致の程度に関する項

5. 上記工程 a では、上記補正対象のパラメータについて最適解探索問題として上記評価関数の値を最適化する際に動的計画法のアルゴリズムを用いることを特徴とする請求の範囲第 3 項記載の試料解析方法。

6. 上記動的計画法のアルゴリズムでは、補正の対象となるパラメータに含

まれるデータ点の最適な対応関係を、スコアを算出して評価する際に、標準物質に由来するデータ点に関する対応関係については、当該スコアを良くする設定とすることを特徴とする請求の範囲第5項記載の試料解析方法。

7. 上記動的計画法のアルゴリズムでは、補正対象のパラメータに含まれる

5 データ点の最適な対応関係を、スコアを算出して評価する際に、標準物質に由来するデータ点に関する対応関係については、必ず指定された点で対応することを拘束条件とすることを特徴とする請求の範囲第5項記載の試料解析方法。

8. 上記試料は、蛋白質群及び／又はペプチド群を含むことを特徴とする請求の範囲第1項記載の試料解析方法。

10 9. 上記複数の試料は、標準物質を含むことを特徴とする請求の範囲第1項記載の試料解析方法。

10 10. 上記標準物質は、ペプチドT (Ala-Ser-Thr-Thr-Asn-Tyr-Thr)、 β カンモルフィン7 (Tyr-Pro-Phe-Pro-Gly-Pro-Ile) 及びこれらの構造類似体からなる群から選ばれる少なくとも1種以上のペプチドであることを特徴とする請求の範囲
15 第9項記載の試料解析方法。

11. 上記標準物質をゲル内に固定した状態で上記試料に添加することを特徴とする請求の範囲第9項記載の試料解析方法。

12. 試料の分析の結果として得られた多次元データを入力する手順aと、

0 入力された多次元データのうち少なくとも1次元のパラメータについてデータ
を補正する手順bと、

上記手順bで補正した後のデータを含む多次元データを、複数の試料について比較する手順cとをコンピュータに実行させる試料解析プログラム。

13. 上記多次元スペクトルデータは、クロマトグラフィー質量分析の結果として得られる、質量／電荷比を示すパラメータと、イオン強度を示すパラメータと、保持時間を示すパラメータとからなる3次元データであり、上記手順bでは保持時間を示すパラメータについてデータを補正することを特徴とする請求の範囲第12項記載の試料解析プログラム。

14. 上記手順bでは、補正対象のパラメータを除いたパラメータに関するプロファイルを参照プロファイルとし、複数の試料間で複数の参照プロファイル

に関する配置類似性の尺度となる評価関数を用いて、当該評価関数の値を最適化する最適解探索問題として各プロファイルの配置を求める特徴とする請求の範囲第12項記載の試料解析プログラム。

15. 上記評価関数は、以下の①～⑥からなる群から選ばれる1以上の項で

5 定義されることを特徴とする請求の範囲第14項記載の試料解析プログラム。

①補正対象のパラメータに関するプロファイル間の類似性及び/又は距離に関する項

②参照プロファイルに関するプロファイル間の類似性及び/又は距離に関する項

③比較対象となるプロファイル間でのデータ点の一致の程度に関する項

10 ④比較対象となるプロファイル間でのデータ点の不一致の程度に関する項

⑤比較対象となるプロファイル間での標準物質由来シグナルの一致又は不一致の程度に関する項

⑥比較操作を繰り返す際に、過去の比較における一致の程度に関する項

16. 上記手順aでは、上記補正対象のパラメータについて最適解探索問題

15 として上記評価関数の値を最適化する際に動的計画法のアルゴリズムを用いることを特徴とする請求の範囲第14項記載の試料解析プログラム。

17. 上記動的計画法のアルゴリズムでは、補正の対象となるパラメータに含まれるデータ点の最適な対応関係を、スコアを算出して評価する際に、標準物質に由来するデータ点に関する対応関係については、当該スコアを良くする設定

0 とすることを特徴とする請求の範囲第16項記載の試料解析プログラム。

18. 上記動的計画法のアルゴリズムでは、補正対象のパラメータに含まれるデータ点の最適な対応関係を、スコアを算出して評価する際に、標準物質に由来するデータ点に関する対応関係については、必ず指定された点で対応することを拘束条件とすることを特徴とする請求の範囲第16項記載の試料解析プログラム。

19. 上記試料は、蛋白質群及び/又はペプチド群を含み、当該蛋白質群及び/又はペプチド群に由来する多次元データを解析することを特徴とする請求の範囲第12項記載の試料解析プログラム。

20. 上記複数の試料は、標準物質を含み、上記手順bではこれら標準物質

に由来する多次元データ及び上記試料に含まれる成分に由来する多次元データを用いることを特徴とする請求の範囲第12項記載の試料解析プログラム。

21. 上記標準物質は、ペプチドT (Ala-Ser-Thr-Thr-Asn-Tyr-Thr)、 β カツモルフィン7 (Tyr-Pro-Phe-Pro-Gly-Pro-Ile) 及びこれらの構造類似体からなる
5 群から選ばれる少なくとも1種以上のペプチドであることを特徴とする請求の範囲第20項記載の試料解析プログラム。

22. 上記標準物質をゲル内に固定した状態で上記試料に添加することを特徴とする請求の範囲第20項記載の試料解析プログラム。

図1

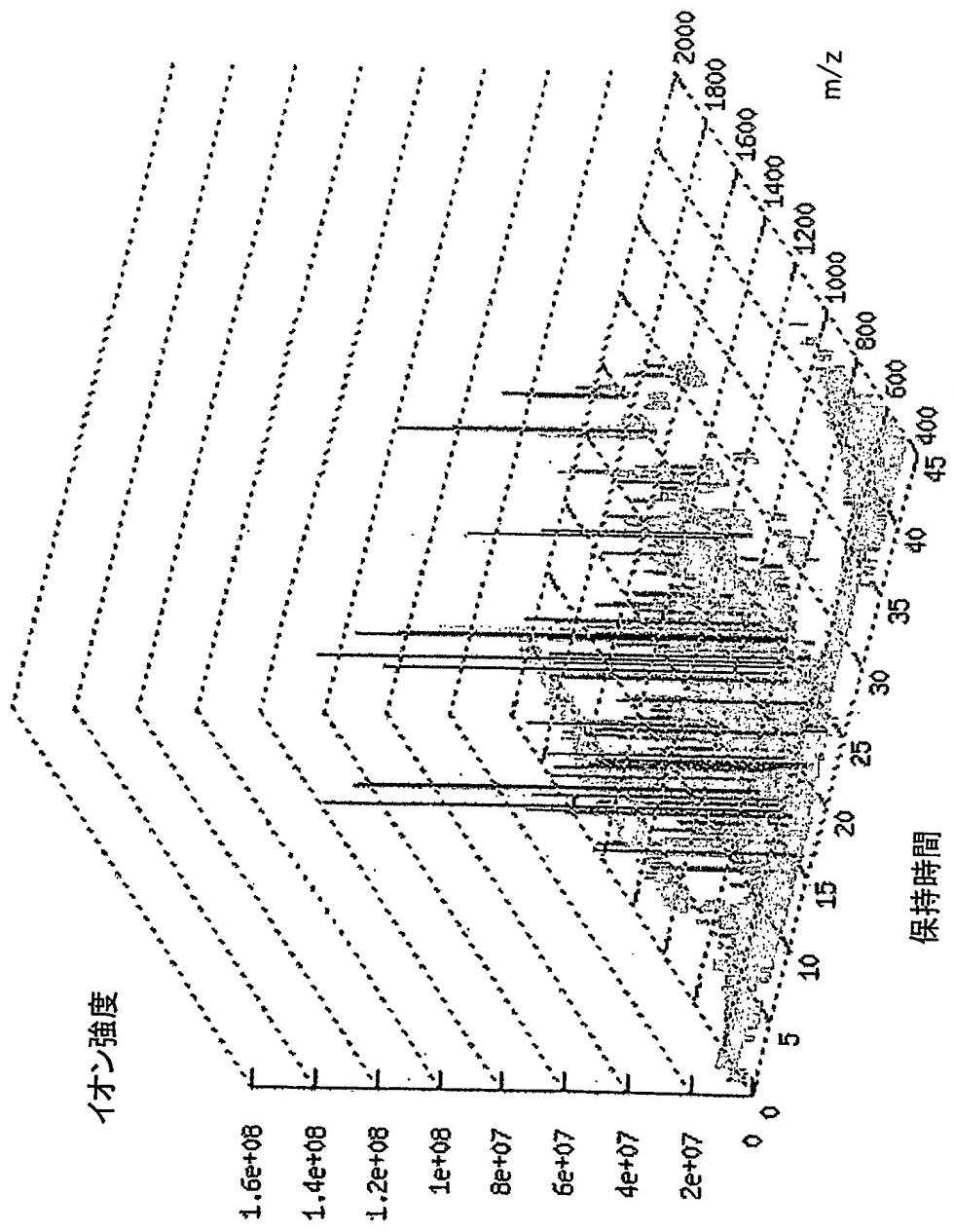


図2

<i>m/z</i>	保持時間	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	..
553						223	121													
560													259	153						
563		564	882	96																
590									98	216										
612																				
613											21	332	48							
:															436	842	1291	461		

m/z	保持時間	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	..
553																				
560																				
563																				
590																				
612																				
613																				
:																				

図 3

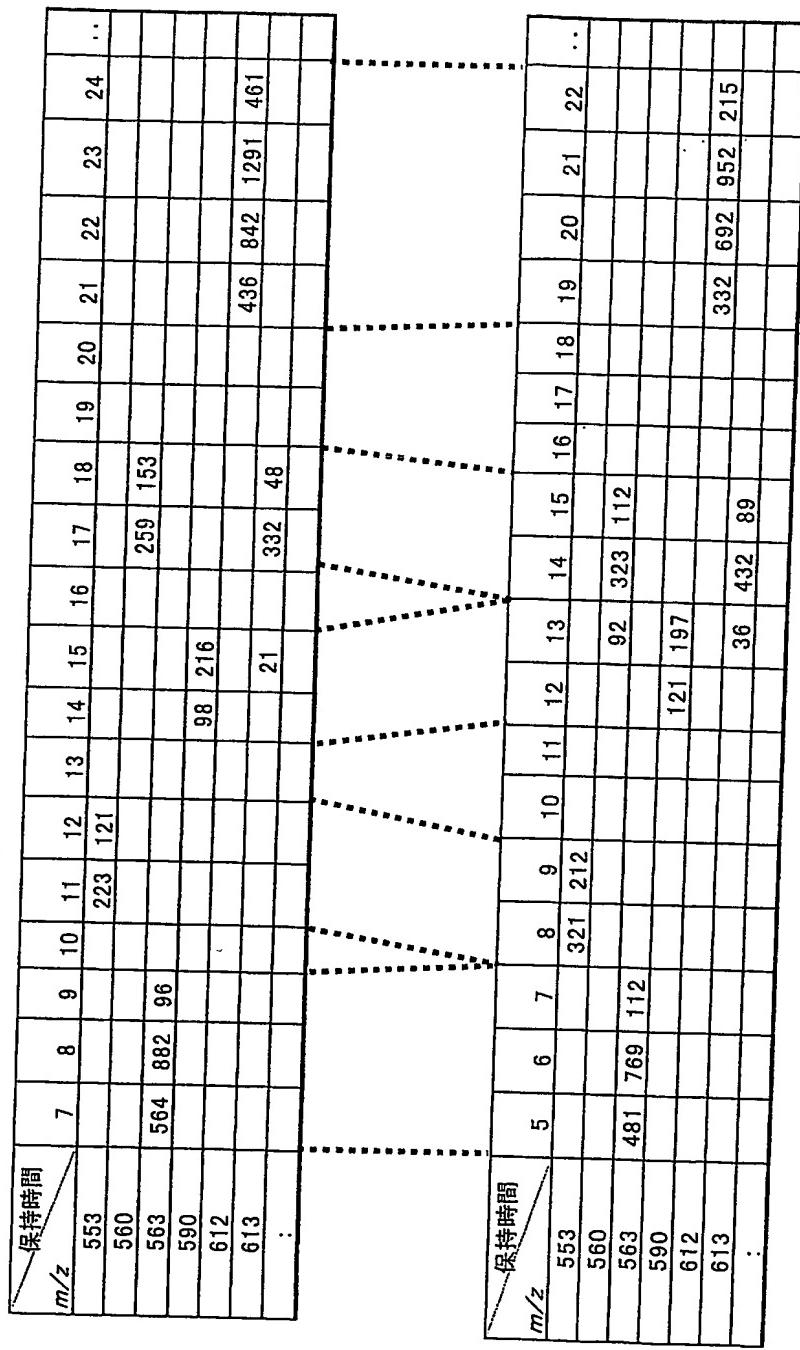


図 4

	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
5																		
6																		
7																		
8																		
9																		
10																		
11																		
12																		
13																		
14																		
15																		
16																		
17																		
18																		
19																		
20																		
21																		
22																		

図 5

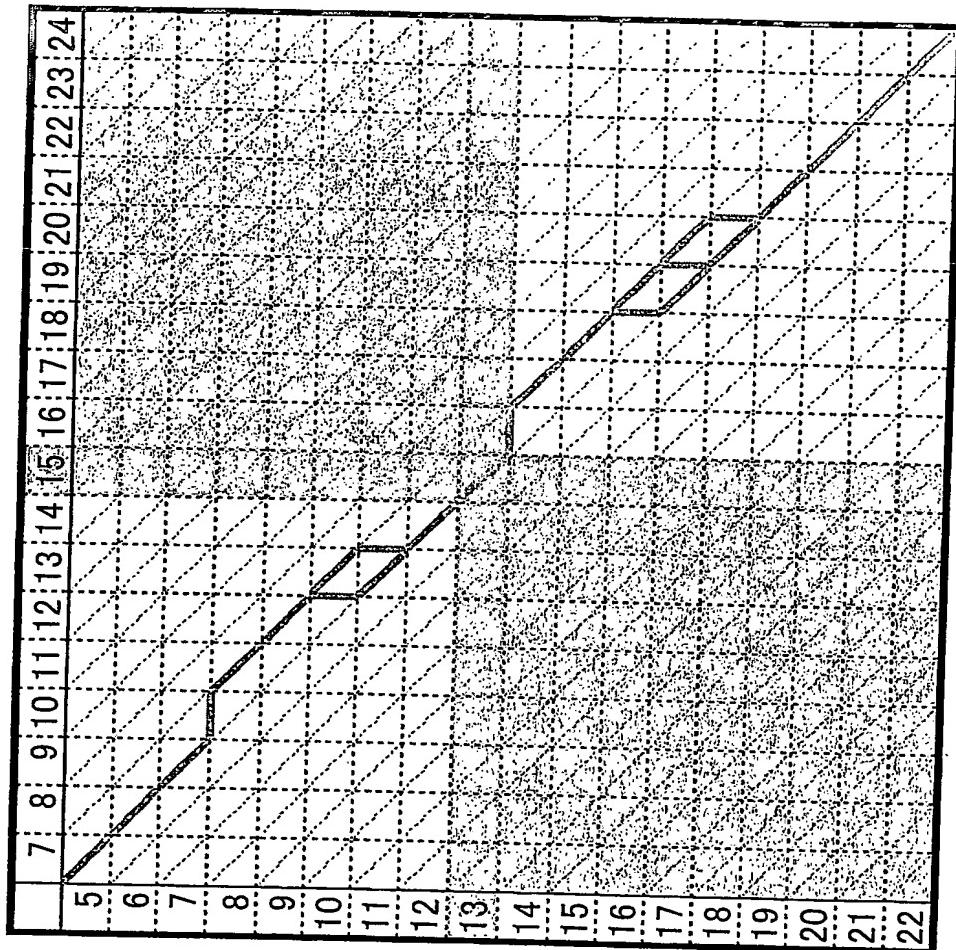


図 6

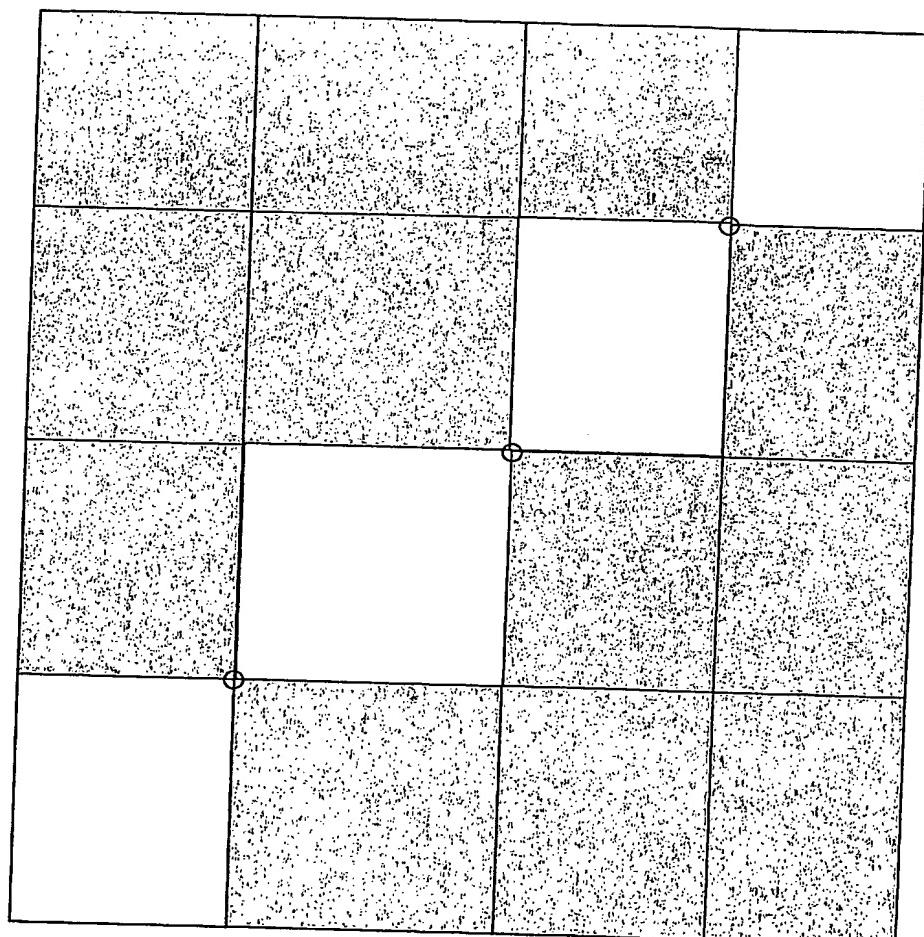


図 7

図 8

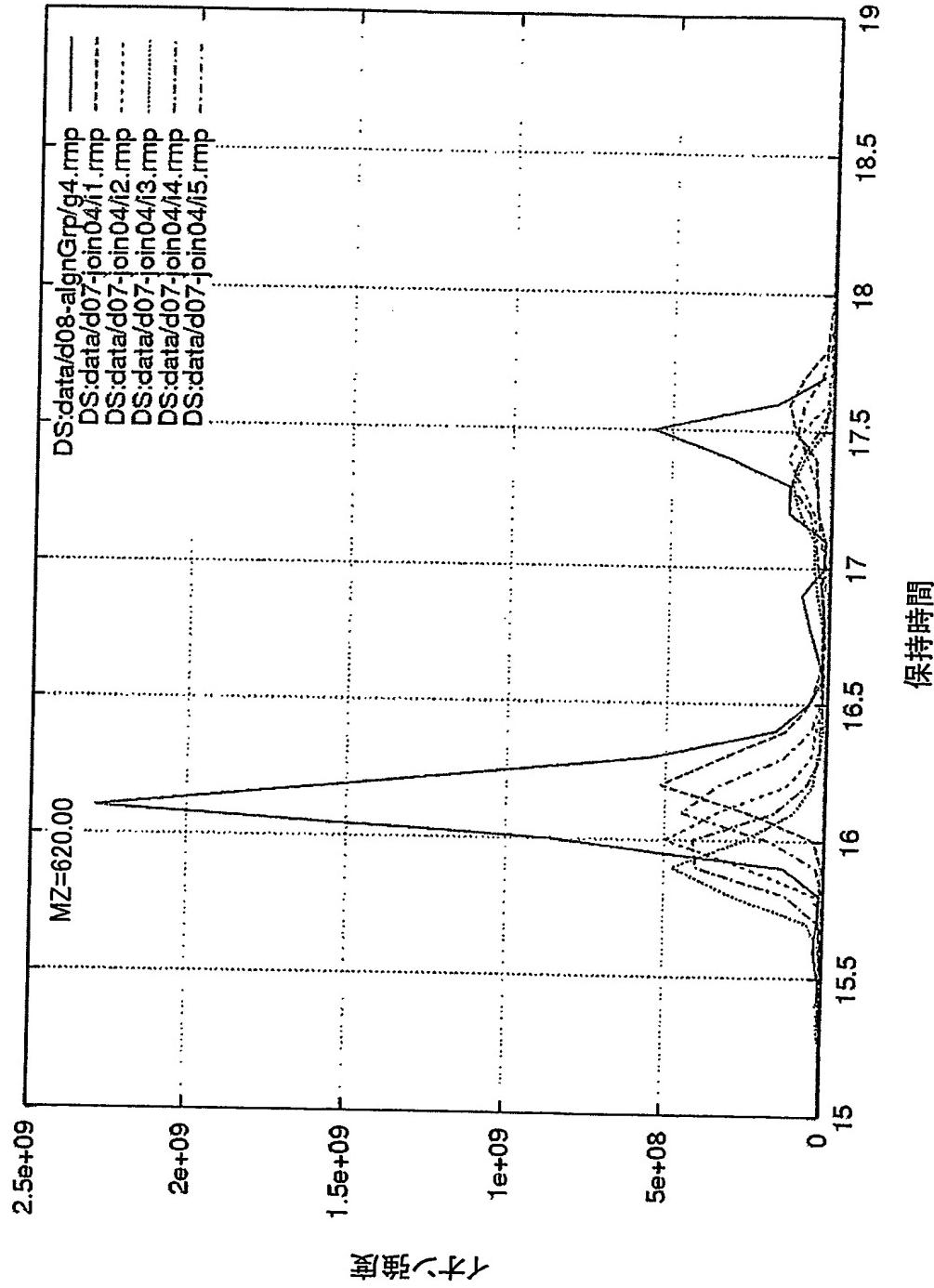


図 9

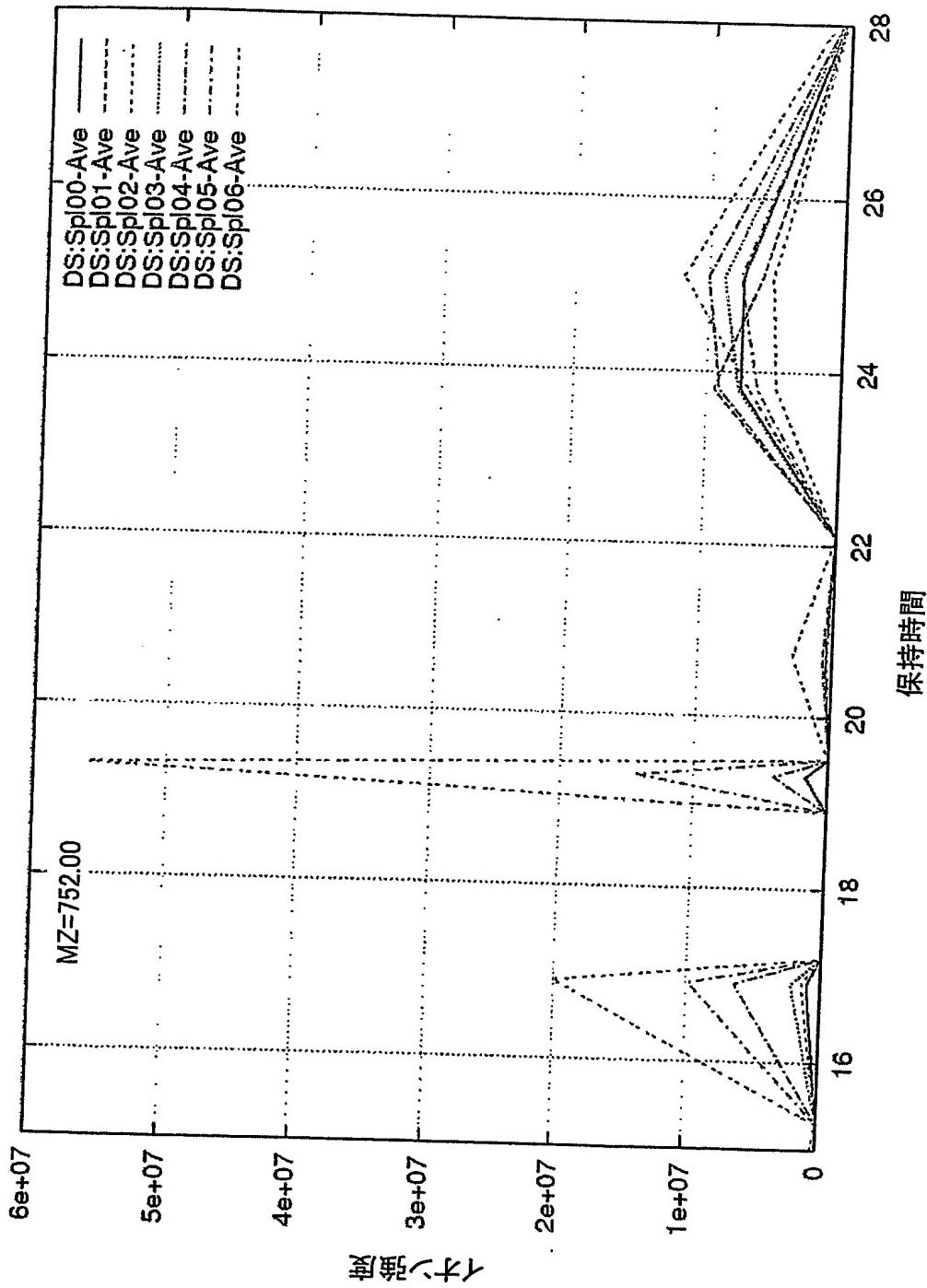


図 10

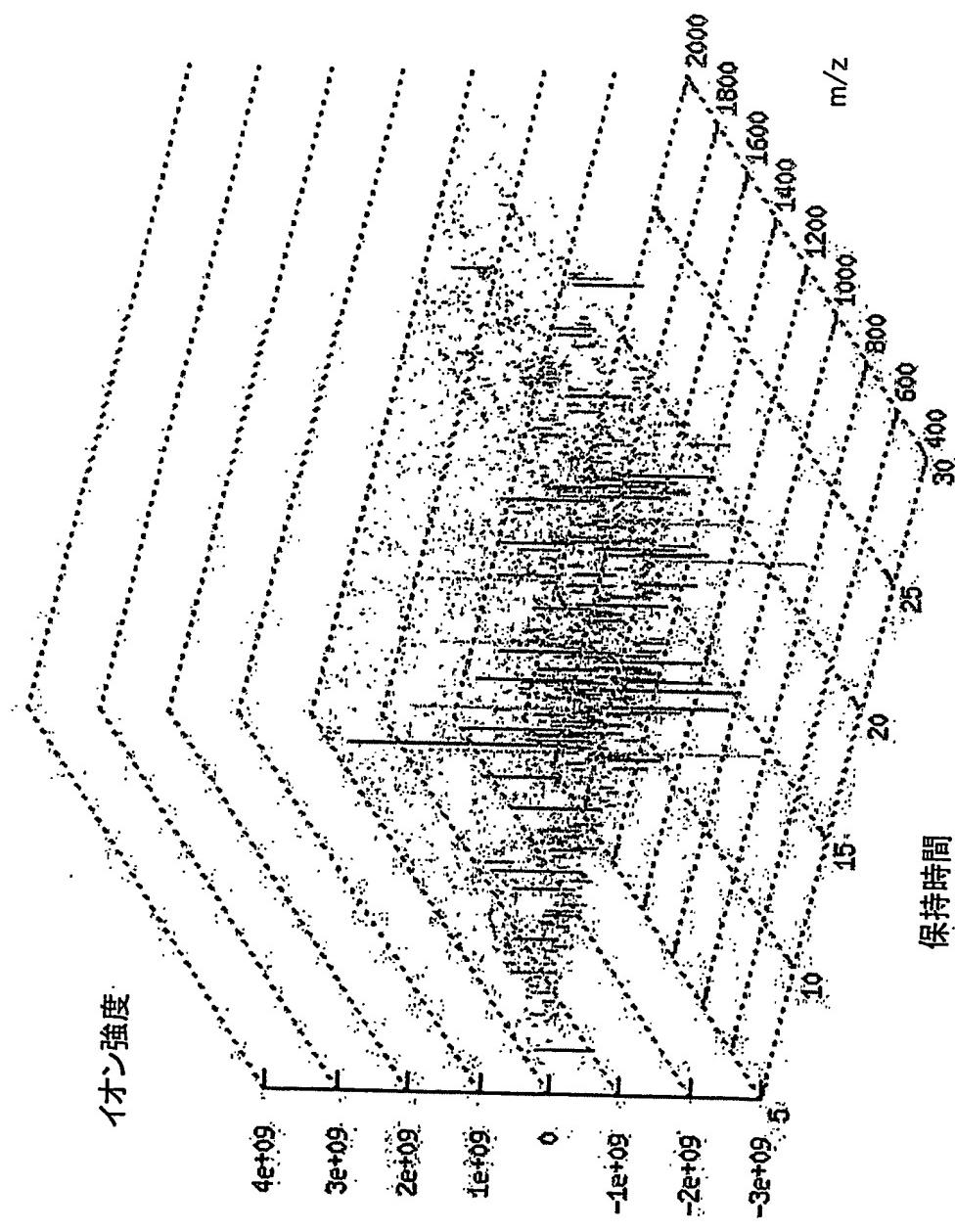


図11

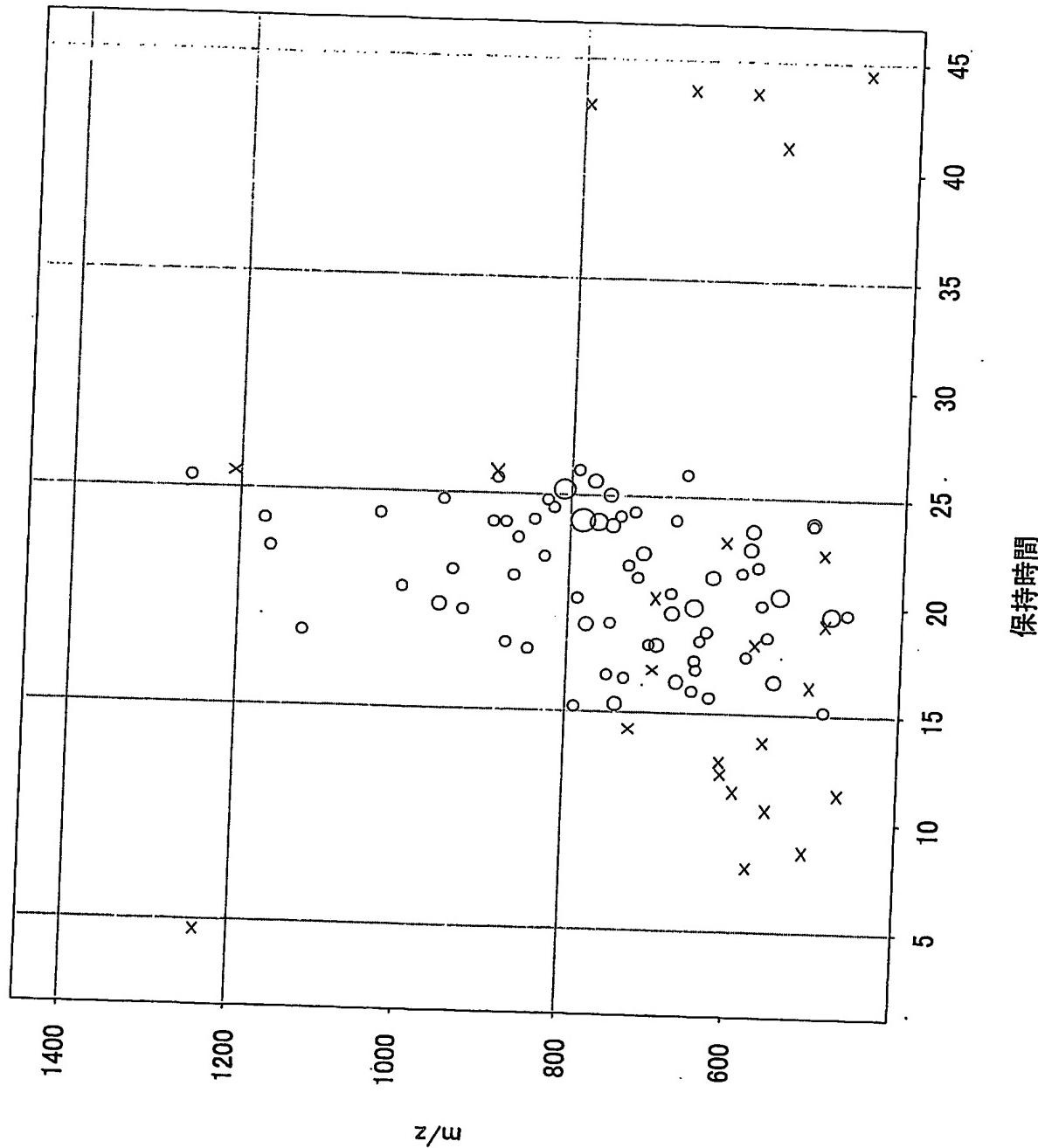


図 1 2

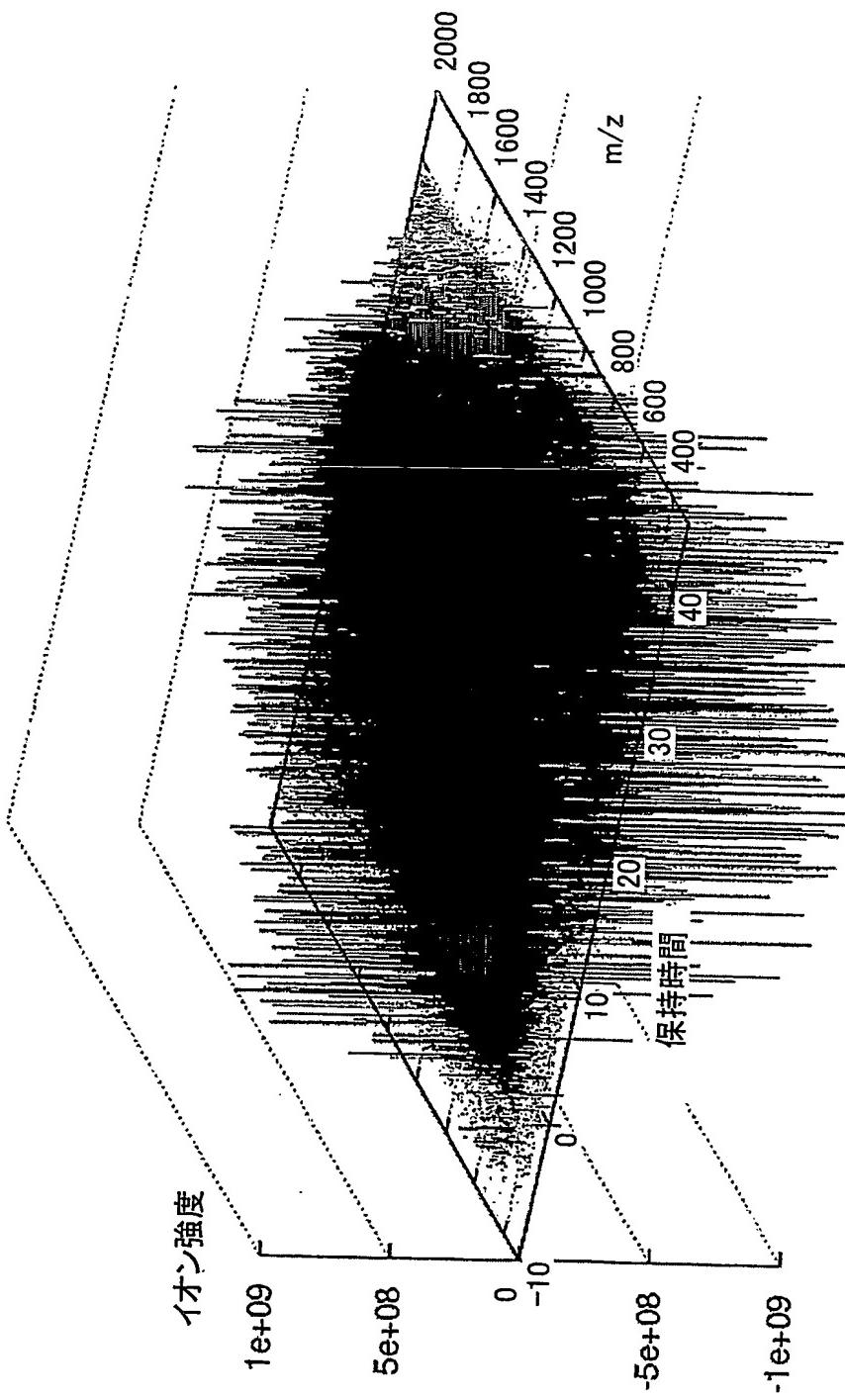


図 13

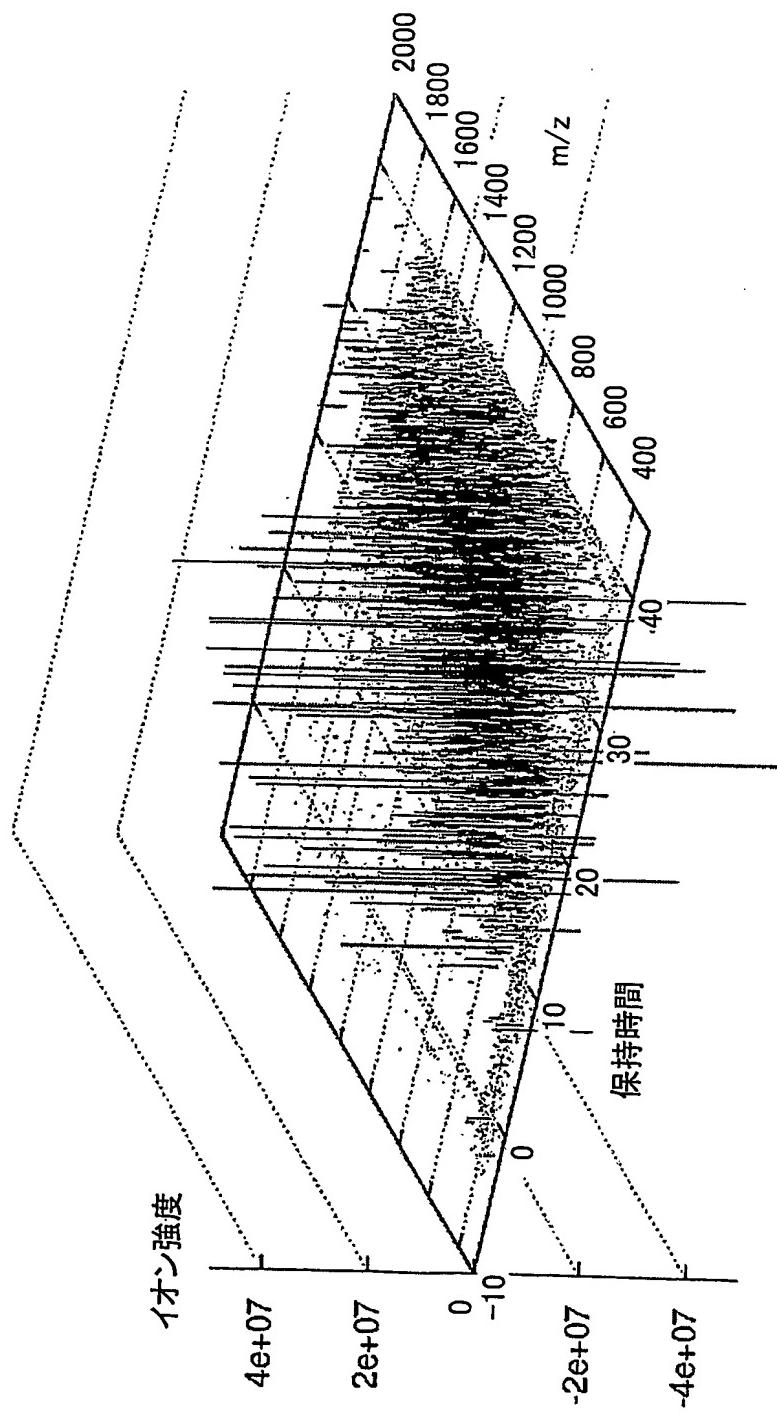


図 14

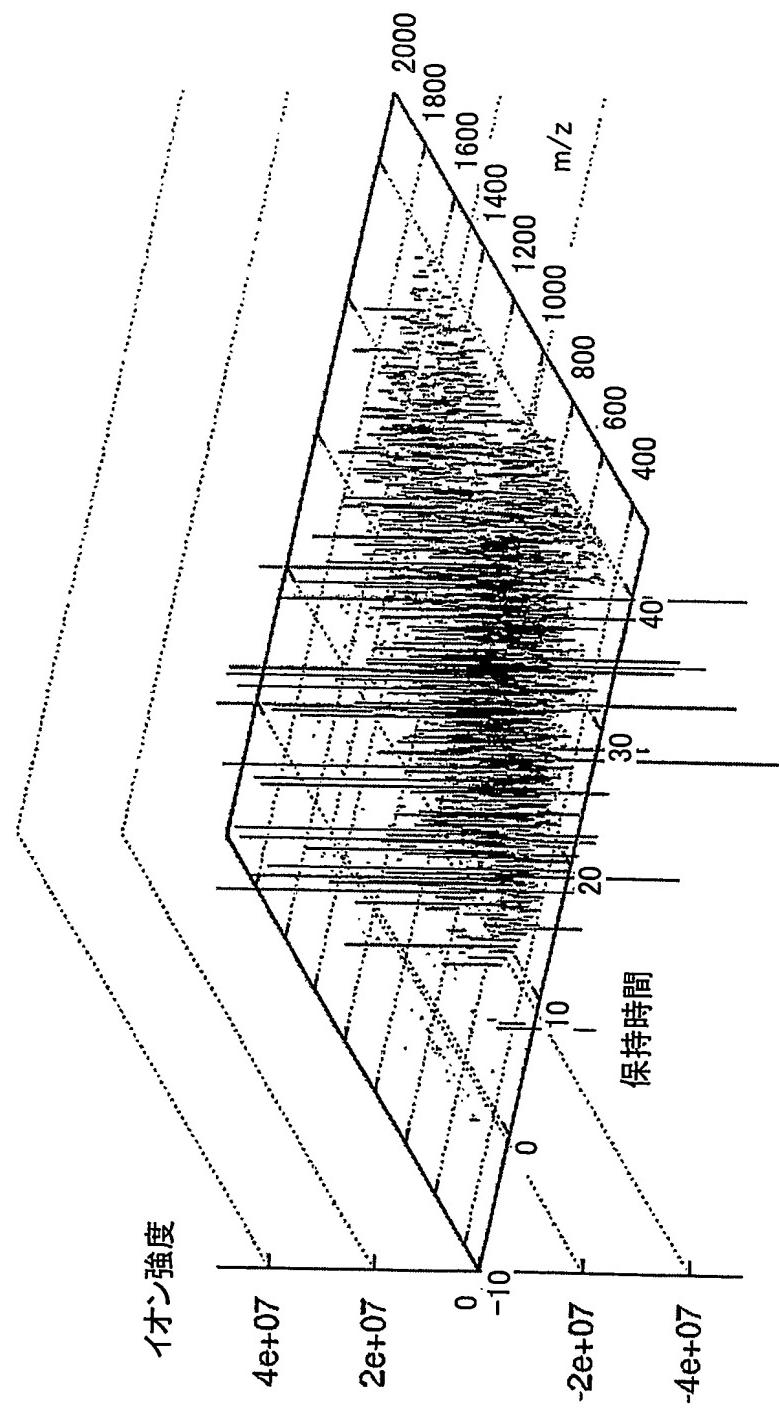


図 15

プロテオームパターン	有意差	蛋白質名	SwissProt
N+特異的	あり	Metastasis associated protein (MTA3)	MTA3_HUMAN
ややN+特異的	顯著	Calcyclin	S106_HUMAN
ややN+特異的	なし	Catenin δ-1	CTD1_HUMAN
N-特異的	なし	Catenin α-1	CTN1_HUMAN
ややN+特異的	顯著	Calmodulin	CALM_HUMAN
ややN+特異的	なし	Calcium/calmodulin-dependent serine protein kinase / hCASK	CSKP_HUMAN
ややN+特異的	顯著	Neuromodulin / Calmodulin-binding protein P-57.	NEUM_HUMAN
ややN+特異的	顯著	Collagen α3(IV)	CA34_HUMAN
ほぼ同量	あり	Collagen α5(IV)	CA54_HUMAN
ややN+特異的	なし	Neutrophil gelatinase-associated lipocalin (NGAL)	NGAL_HUMAN
ややN+特異的	あり	Fibronectin	FINC_HUMAN
ほぼ同量	あり	A disintegrin and metalloproteinase with thrombospondin motifs 15 (ADAMTS-15)	AT15_HUMAN
ややN+特異的	あり	A disintegrin and metalloproteinase with thrombospondin motifs 19 (ADAMTS-19)	AT19_HUMAN
ややN+特異的	あり	A disintegrin and metalloproteinase with thrombospondin motifs 2 (ADAMTS-2)	AT2_HUMAN
ややN+特異的	あり	Integrin α-3 / VLA-3 α	ITA3_HUMAN
ややN+特異的	あり	Integrin α-6 / VLA-6	ITA6_HUMAN
ややN+特異的	あり	Integrin α-11.	ITAH_HUMAN
N+特異的	顯著	Integrin α-M / CD11b / Leukocyte adhesion receptor MO1	ITAM_HUMAN
ややN+特異的	あり	Integrin β-1 / VLA-4 β	ITB1_HUMAN
N+特異的	顯著	Laminin α-2	LMA2_HUMAN
ややN+特異的	あり	Laminin α-4	LMA4_HUMAN
ややN+特異的	あり	Laminin γ-1	LMG1_HUMAN
ややN+特異的	あり	Laminin γ-2	LMG2_HUMAN
ややN+特異的	顯著	Matrin 3.	MAT3_HUMAN
ややN+特異的	あり	Nucleophosmin (NPM) / Numatrin	NPM_HUMAN
ややN+特異的	顯著	Tenascin	TENA_HUMAN
ややN+特異的	あり	Tissue inhibitor of metalloproteinases-3 (TIMP-3)	TIM3_HUMAN
ややN+特異的	顯著	Urokinase plasminogen activator surface receptor (uPAR)	UPAR_HUMAN
N+特異的	あり	Vinculin	VINC_HUMAN
ややN+特異的	あり	TIE-2	TIE2_HUMAN
ややN+特異的	顯著	Insulin-like growth factor binding protein complex acid labile chain (ALS)	ALS_HUMAN
N+特異的	あり	EGF	EGF_HUMAN
ややN+特異的	顯著	EGFR kinase substrate EPS8	EPS8_HUMAN
N+特異的	あり	Insulin-like growth factor binding protein 2 (IGFBP-2)	IBP2_HUMAN
ほぼ同量	顯著	Mast/stem cell growth factor receptor (SCFR)	KIT_HUMAN
N+特異的	あり	β-nerve growth factor (β-NGF).	NGF_HUMAN
ややN+特異的	あり	VEGFR-3	VGR3_HUMAN
N+特異的	あり	EGFR / ErbB-1	EGFR_HUMAN
ややN+特異的	あり	ErbB-2 / HER2	ERB2_HUMAN
ほぼ同量	なし	ErbB-3 / HER3	ERB3_HUMAN
N+特異的	あり	MEKK3	M3K3_HUMAN
ややN+特異的	あり	MEKKK6	M4K6_HUMAN
ややN+特異的	なし	MAPKK7 / JNKK2	MPK7_HUMAN
ややN+特異的	なし	HSP75 / TRAP-1	TRAL_HUMAN
ややN+特異的	あり	Nucleoside diphosphate kinase B (NDK B) / nm23-H2	NDKB_HUMAN
N+特異的	顯著	Serine/threonine-protein kinase PAK 1 / p21-activated kinase 1	PAK1_HUMAN
ややN+特異的	なし	Interferon-regulated resistance GTP-binding protein MxA / IFI-78K	MX1_HUMAN
ややN+特異的	あり	Transcriptional coactivator Sp110 / Interferon-induced protein 41/75	SP11_HUMAN
ややN+特異的	あり	Interleukin-12 α (IL-12A)	I12A_HUMAN
ややN+特異的	なし	Interleukin 18 receptor 1	IR18_HUMAN
N+特異的	あり	Cytochrome P450 3A43	C343_HUMAN
N+特異的	あり	Cytochrome P450 3A7	CP37_HUMAN
ややN+特異的	顯著	Cytochrome P450 4F3	CPF3_HUMAN

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/004621

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl⁷ G01N27/62, G01N30/72

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl⁷ G01N27/62-27/70, G01N30/72Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-2004
Kokai Jitsuyo Shinan Koho 1971-2004 Jitsuyo Shinan Toroku Koho 1996-2004Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JICST FILE (JOIS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 7-198703 A (Tokai Rubber Industries, Ltd.), 01 August, 1995 (01.08.95), Full text; Figs. 1 to 12 (Family: none)	1,12
Y	J.D.Pinkston, et al., "Characterization of low molecular weight alkoxylated polymers using long column SFC/MS and an image analysis based quantitation approach", JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY Vol.13, No.10, October, 2002, pages 1195 to 1208	2,13
A	JP 63-108260 A (JEOL Ltd.), 13 May, 1988 (13.05.88), Full text; Figs. 1 to 5 (Family: none)	1-2,12-13

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	
"A"	document defining the general state of the art which is not considered to be of particular relevance
"E"	earlier application or patent but published on or after the international filing date
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O"	document referring to an oral disclosure, use, exhibition or other means
"P"	document published prior to the international filing date but later than the priority date claimed
"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&"	document member of the same patent family

Date of the actual completion of the international search
09 August, 2004 (09.08.04)Date of mailing of the international search report
24 August, 2004 (24.08.04)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/004621

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 11-344482 A (JEOL Ltd.), 14 December, 1999 (14.12.99), Full text; Figs. 1 to 7. (Family: none)	1-2, 12-13

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/004621

Box No. II**Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III**Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

The technical feature common to claims 1, 2; 12, 13, claims 3-7, 14-18, claims 8, 19, and claims 9-11, 20-22 is that at least a one-dimensional parameter, out of multi-dimensional data collected by analyzing a sample, is corrected, and the corrected data is compared with data on other samples. However, such processing of analysis data is not novel since it is disclosed in documents such as JP 7-198703A (Tokai Rubber Industries, Ltd.), 1 August, 1995 (01.08.95). Consequently, this common technical feature is not a special technical feature within the meaning of PCT Rule 13.2, second sentence, since it makes no contribution over the prior art.

(Continued to extra sheet.)

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos. 1, 2; 12, 13

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/004621

Continuation of Box No.III of continuation of first sheet(2)

Since there exists no other common feature which can be considered as a special technical feature, no technical relationship within the meaning of PCT Rule 13 between the different inventions can be seen. Consequently, claims 1, 2; 12, 13, claims 3-7, 14-18, claims 8, 19, and claims 9-11, 20-22 do not satisfy the requirement of unity of invention.

国際調査報告

国際出願番号 PCT/JP2004/004621

A. 発明の属する分野の分類(国際特許分類(IPC))

Int. C17

G01N27/62, G01N30/72

B. 調査を行った分野

調査を行った最小限資料(国際特許分類(IPC))

Int. C17

G01N27/62-27/70, G01N30/72

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2004年
日本国登録実用新案公報	1994-2004年
日本国実用新案登録公報	1996-2004年

国際調査で使用した電子データベース(データベースの名称、調査に使用した用語)

JICSTファイル(JOIS)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP 7-198703 A(東海ゴム工業株式会社) 1995.08.01	1, 12
Y	全文、第1-12図(ファミリーなし)	2, 13
Y	J.D.Pinkston, et al.: "Characterization of low molecular weight alkoxylated polymers using long column SFC/MS and an image analysis based quantitation approach" JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY Vol. 13, No. 10, October 2002, pp. 1195-1208	2, 13

 C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

- 「A」特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献(理由を付す)
 「O」口頭による開示、使用、展示等に言及する文献
 「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

- 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」同一パテントファミリー文献

国際調査を完了した日

2004.08.09

国際調査報告の発送日 24.8.2004

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官(権限のある職員)
高場 正光

2W 2910

電話番号 03-3581-1101 内線 3290

C(続き) .	関連すると認められる文献	
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 63-108260 A (日本電子株式会社) 1988. 05. 13 全文, 第1—5図 (ファミリーなし)	1-2, 12-13
A	JP 11-344482 A (日本電子株式会社) 1999. 12. 14 全文, 第1—7図 (ファミリーなし)	1-2, 12-13

第II欄 請求の範囲の一部の調査ができないときの意見（第1ページの2の続き）

法第8条第3項（PCT17条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. 請求の範囲 _____ は、この国際調査機関が調査をすることを要しない対象に係るものである。つまり、
2. 請求の範囲 _____ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. 請求の範囲 _____ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

第III欄 発明の単一性が欠如しているときの意見（第1ページの3の続き）

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

請求の範囲1-2, 12-13, 請求の範囲3-7, 14-18, 請求の範囲8, 19, 請求の範囲9-11, 20-22に共通の事項は、試料の分析の結果として得られた多次元データのうち少なくとも1次元のパラメータを補正し、補正した後データを複数の試料について比較すること、である。しかしながら、このような分析データの取り扱いは、例えば文献JP 7-198703 A（東海ゴム工業株式会社）1995.08.01に開示されているから新規でなく、先行技術の域を出ないから、PCT規則13.2の第2文の意味において、この共通事項はない。

特別な技術的特徴と考えられる他の共通の事項は存在しないので、それらの相違する発明の間にPCT規則13の意味における技術的な関連を見いだすことはできない。よって、請求の範囲1-2, 12-13, 請求の範囲3-7, 14-18, 請求の範囲8, 19, 請求の範囲9-11, 20-22は発明の単一性を満たしていない。

1. 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。
2. 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。
3. 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったので、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。
4. 出願人が必要な追加調査手数料を期間内に納付しなかったので、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

請求の範囲 1-2, 12-13

追加調査手数料の異議の申立てに関する注意

- 追加調査手数料の納付と共に出願人から異議申立てがあった。
- 追加調査手数料の納付と共に出願人から異議申立てがなかった。